A DECISION MODEL FOR WORK DISTRIBUTION IN GLOBAL SOFTWARE DEVELOPMENT

BY

ABDULRHMAN AHMED ALSRI

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

SOFTWARE ENGINEERING

SEPTEMBER, 2015

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **ABDULRHMAN AHMED ALSRI** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN SOFTWARE ENGINEERING**.

Thesis Committee

Dr. Sajjad Mahmood (Adviser)

(Co-adviser)

Dr. Sultan Al-Muhammadi (Member)

Dr. Mahmood Niazi (Member)

(Member)

Dr. Abdulaziz Alkhoraidly
Department Chairman

Dr. Salam A. Zummo
Dean of Graduate Studies

29/12/15
Date

Dedication

To my mother, my father, my wife, my sons and daughters, my brothers and sisters and my best friends.

ACKNOWLEDGMENTS

Finally, I acknowledge KFUPM and Information and Computer Science Department and faculty members who taught me.*Thanks and all praise is to Allah who created me, blessed me with health and granted me success.*

I state my deep appreciation and thanks to my advisor Dr. Sajjad Mahmood who has always been providing kind support, continuous encouragement and supervision.

I thank my committee members, Dr. Sultan Al-Muhammadi and Dr. Mahmood Niazi for their valuable advice and guidelines.

And, I would like to thank my scholarship sponsor, Hadhramout Establishment for Human Development.

Finally, I acknowledge KFUPM and Information and Computer Science Department and faculty members who taught me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GSD Global Software Development

ANN Artificial Neural Networks

KNN K-Nearest Neighbors

THESIS ABSTRACT

NAME: Abdulrhman Ahmed Alsri

TITLE OF STUDY: **A decision model for work distribution in global software development**

MAJOR FIELD: Software Engineering

DATE OF DEGREE: September, 2015

Global Software Development (GSD) initiative aims to facilitate software development process by providing access to skilled workers at a relatively low cost and 24/7 software development model. Previous work suggests that half of the companies that have tried GSD have failed to realize the anticipated outcomes which have resulted in poor outsourcing relationships, high costs and overall poor software products. One critical factor for successful GSD projects is the allocation of tasks as project managers not only need to consider their workforce but also need to take into the account the characteristics of the geographically distributed sites and their relationships. This research introduces a model for global work distribution based on machine learning techniques, that considers multiple GSD factors

that affect the decision. The model outputs, for project managers, the suitable site for a given tasks.

ملخص الرسالة

الإسم الكامل: عبدالرحمن أحمد محسن السري

عنوان الرسالة: أداة لإتخاذ القرار لتوزيع عمل تطوير أنظمة الحاسوب عالمياً

التخصص: هندسة البرمجيات

تاريخ الدرجه العلمية: سبتمبر 2015

تطوير أنظمة الحاسوب عالمياً تسهل عملية إنتاج أنظمة الحاسوب من خلال إتاحة الوصول إلى خبراء مؤهلين بتكاليف منخفضه وإتاحة العمل على مدار 24 ساعه لمدة 7 أيام في الإسبوع. الإبحاث السابقة أخبرت بأن نصف الشركات تقريباً التي إستخدمت مفهوم تطوير أنظمة الحاسوب عالمياً فشلت لتحقيق الفوائد المتوقعه مما أدى الى إستيراد خدمات ضعيفه وبأسعار مرتفعه وبجودة أقل للأنظمة التي تم إنتاجها. ثم إن من أهم العوامل لنجاح مشروع تطوير أنظمة الحاسوب عالمياً هو ليس الأخذ بعين الإعتبار التكاليف الماليه فقط, وإنما يجب كذلك الأخذ بعين الإعتبار الخصائص الجغرافيه والعلاقه فيما بينها للأماكن المتفرقه التي سيتم توزيع العمل إليها. هذا البحث يقدم أداة لإتخاذ قرار قائم على تقنيات تعلم الآله أو الذكاء الإصطناعي والتي تأخذ بعين الإعتبار العوامل المؤثره على قرار توزيع عمل تطوير أنظمة الحاسوب عالمياً. الأداة تقدم حلاً لمدراء مشاريع تطوير أنظمة الحاسوب عالمياً والأماكن المناسبة لإسناد العمل إليها.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Global Software Development (GSD) is a paradigm that is defined as a company (client) contracts all/part of its software development's activities to another company (vendor), who provides services in return for remuneration [2]. Over the last decade and half, many firms in the world have started GSD projects. The approximate annual volume of GSD in 2010 was $100 billion [1]. A large number of these organizations have adopted GSD to reduce development cost and increase overall software quality [3, 4]. Organizations benefit from GSD because vendors in developing countries cost less than in-house development [5]. Furthermore, client organizations also get access to skillful human workforce and take advantage of the 24/7 development time model.

Despite steady growth of GSD projects, half of the companies that have tried GSD have failed to realize the anticipated outcomes which has resulted in poor global relationships, misunderstanding of the projects' requirements and higher development costs [6]. There are many reasons for these failures [7, 8, 9]. These failures are usually traced back to two main causes: insufficient abilities at remote sites such as absence of domain knowledge, high turnover rate etc., and problems at the interfaces between distributed sites due to communication, coordination and cultural barriers [4].

Work distribution or task allocation is considered to be a critical step in GSD [9, 10, 11]. At that step, we decide the sites or GSD vendors that are suitable for our development tasks.

In this research, we manipulate multiple factors that affect task allocation. We present a systematic decision making method for task allocation that facilitates the success in offshore development and avoid additional efforts in adopting the GSD concept. The method helps in distinguishing a fit site for a given task. However, this model is validated using hypothetical data.

1.2 Research Problem Statement

Despite the benefits of Global Software (GSD) Development, some researches argue that the expected benefits of GSD is decreasing because of the required overhead of communication and coordination between teams [10]. For example, Aron, Ravi et al. [12] sated that half of organizations which shift processes offshore fail to generate estimated financial outcomes, and some of them fail completely [13]. That is because cost-benefit analysis is not fully understood [14]. However, although the existence of these drawbacks, offshore development is rather increasing in the software industry [15].

In GSD prject, task allocation decision is a critical step because project managers have to be familiar with vendor site capabilities, time zone differences and developing skills [16]. For example, researchers studies discuss that allocating tasks to different time-zone locations maximizes the working hours that help in achieving tasks quickly and increase productivity [17, 18]. Furthermore, if an organization does not have sufficient skills in its project staff at certain phase, it might be the time for it to access more skilled people globally. Other organizations do offshoring development in order to look for lower cost rates of development [1].

One critical factor for successful GSD projects is the allocation of tasks as project managers not only need to consider their workforce but also need to take into the account the characteristics of the geographically distributed sites and their relationships. For that reason, we need to have a systematic task allocation process in order to decide the best task allocation for our project work. A number of GSD challenges are directly impacted by the decisions taken at the task allocation phase of GSD projects [19]. Despite the importance of this problem, little research has been carried out to improve task allocation processes at GSD organizations.

1.3 Research Objectives

In this research, a task allocation model is presented that predicts the fit sites for a given set of tasks based on ANN and KNN algorithms. After that, we evaluate the performance of the two algorithms. Also, the Simplex algorithm is used for assigning the given tasks to the fit predicted sites. In this thesis, the research objectives are as follows:

- First, generate hypothetical data that is generated by using some collected-manually real data. This way allow our dataset to reflect kind of real data for our model simulation.

- Second, develop a task allocation model based on machine learning, the Artificial Neural Networks (ANN), to predict the fit site for a given task. The decision model considers multiple GSD factors.

- Third, the model includes another machine learning, K-Nearest Neighbors Classification, that predicts the fit site using different concept.

- Then, the performance of the two techniques, neural networks and K-Nearest neighbor classification is evaluated and compared.

- After that, the Simplex algorithm is sued to assign the tasks automatically to the predicted sites.
- Finally, the model is validated using a real case study as a part of the model validation.

1.4 Outlines

The rest of this research proposal is organized as follows: Chapter 2 highlights the overview of global software development and provides literature review focusing on work distribution related work. Chapter 3 presents the research methodology. Chapter 4 presents the case study. Chapter 5 presents Discussion. Chapter 6 discusses Conclusion, Future Work & appendices

CHAPTER 2

BACKGROUND & RELATED STUDIES

In this chapter, we introduce an overview about GSD. After that, the classified literature review is presented.

2.1 UNDERSTANDING GLOBAL SOFTWARE DEVELOPMENT

Global software Development (GSD) is software development and maintenance in globally distributed places. Today, businesses include offshoring as a key element for cost and quality flexibility. Below is a definition of offshoring and outsourcing development which are two views of global development [1]:

- **Offshoring Development:** the business activity beyond sales which takes place outside the country. Enterprise has local branches in low-cost countries.

- **Outsourcing Development:** is a business's relationship with supplier who executes business activities for enterprise outside the enterprise.

2.1.1 Benefits of Global Software Development

There are several benefits that we gain form GSD [1]:

1. **Talent** Computer science skills are very important in the software development. Many countries have no enough skilled people to cooperate in their software products. As a result, firms go globally to find excellent software engineers.

2. **Flexibility** When a software is being developed, many people with broad experience are needed. One step further after the development, we need to outsource people with different skills and levels of experience.

3. **Efficiency and Productivity** Products must be delivered fast because development companies are in a competition with literally a mouse click away. Development companies compete to achieve better cost rates, quality and development time. By using offshore development, we can access to short-time-to profit by using follow the sun and developing and maintaining software in two to three different time zones.

Take a look at Figure 1 for reasons that companies select global development for.

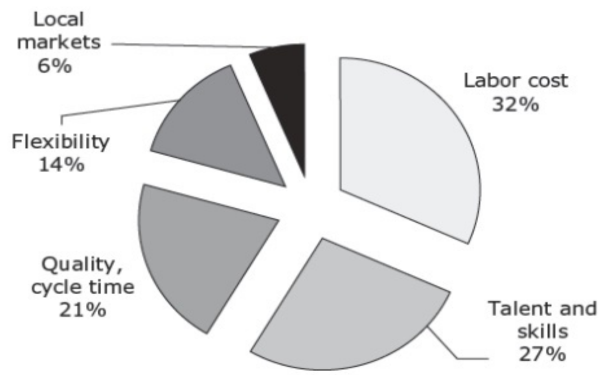

Figure 2.1: Reasons for Outsourcing and Offshoring [1].

2.1.2 Drawbacks of Global Software Development

Working in a global context obviously provides to us many benefits but there are also some drawbacks.

1. **Lack of strategy** Lack of strategy in the work leads to insufficient collaboration and product quality. That might happen if plans do not provide clear work split. Vague strategies might result to insecurity of teams.

2. **Insufficient communication** People are usually motivated to discuss problems when they are close to each other. Teams prefer to talk in informal language and communication [20]. Distributed teams are less effective than a collected team working in the same task.

3. **Inadequate global management** which results in the lack of visibility and task management. Managers usually get scared from lack of management so they tend to split tasks into small ones to be more controlled.

4. **Isolated learning** Sometimes individual sites have their own tools and processes even they are working in the same product lines. They often issue

different work infrastructures in order to differentiate from one another. As a result, low quality support will be fed to the parent organization [1]. Furthermore, distributed development is slower and less forgiving in case of mistakes [21].

5. **Less agility compared with collocated teams** when a task is done, workflow, monitoring, and engineering processes must all be strengthened to make sure requirements are satisfied. Sometimes, if teams are not well-trained they try to escape which causes major trouble during development [22, 23, 24].

2.2 Work Distribution Related Studies

The studies in global software development work distribution has increased since the last decades, addressing the work distribution problems and proposing new solution for that. In this section, selected existing work distribution models are surveyed.

2.2.1 GSD Overhead

Setamanit, et al. [25] used a simulation model to argue that although follow-the-sun concept is reducing development time, but it is rather overhead of coordination. The model focused on site selection decision. However, the model considered dependencies between sites but not between tasks. They developed a method that considers discrete-event and system-dynamic simulation. It concludes information about the impact of different strategies on productivity. However, The drawback of this method is that it can conclude only general statements about task allocation and not a systematic decision.

Doma et al. [26] argues that global development negatively increases the time and cost due to insignificant coordination and communication between geographically distributed sites and people from different cultural backgrounds. The model is based on the critical path calculation for project scheduling and only considers one factor which is time as a task allocation factor.

Researchers at SIMENS company [27] stated some lessons from their GSD experience. Their aim of GSD was to assign tasks to low cost countries. They learned that GSD was rather more complex as a result of the overhead of coordination, changing technologies, and multicultural agreements. They are continuing their research to find a better method for knowledge management in GSD.

2.2.2 Key Factors

Lamersdorf et al. [19] presented an empirical study to identify key factors that impact work distribution of a GSD project. The study identified more than ten factors that affect work distribution decision. The development cost is considered in almost all companies when they plan for task allocation process in global development projects.

Furthermore, Lamersdorf, Ansgar et al. [28] argue that most of todays' development distribution consider only development cost factor. Other factors like workforce capabilities, cultural differences and innovation potential of different regions are not recognized sufficiently. So, they aim towards building multi-criteria distribution model. In their research, they formulate general requirement for work distribution. They finally concluded that no model that support dependencies and relation between sites and tasks, and is a multi-objective supportive in the same time. Also, they concluded that a model fulfilling all the distribution requirements might be too complex and difficult to achieve.

2.2.3 Single Purpose

Another way of distributing the work, by Mockus and Weiss [18], is allocating work items based on their modularity or dependency. That is, allocating group of tasks independently to one site according to the fact that they have dependent code changes. The aim of that is to reduces coordination and communication barriers and to increase productivity. The drawback of that method that it considers only one criteria that is, reducing communication and coordination. Furthermore, the available expertise or the cost rate per site are not considered.

Madachy, Ray [29] extended the COCOMOII model to count the cost for work distribution. It counts the costs based on the personal needs, and the effort at the site according to the site characteristics. However, the other task allocation factors are not considered such as expertise, time zones, etc.

Jabangwe, R. and Smite, D. [30] discussed different stratifies that affect the eventual product quality. However, they only considered the quality factor for decision. Other factors are not fulfilled.

2.2.4 Non Systematic Solutions

Beecham et al. [16] presented a decision support method for global software development that gives a set of recommendations to project managers with an aim to facilitate appropriate site selection that best meets business needs of the project. The drawbacks is that it gives general statements and not systematic decision.

Lately, Wickramaarachchi and Lai [10] presented a task allocation model based on a high abstraction level of the development process models. The model selects the suitable site based on work and site dependency. It identify work-specific characteristics, relation between the work phases, dependency between sites and the site-specific characteristics. The site is selected if its output is maximum value of the four parameters multiplication. The drawback of this method is that each site is processed using manually for each site.

There are other approaches introduced to reduce distance in GSD [17]. First, researchers considered unstructured tasks with difficult-to-use methods and ambiguous output to be more complex coordinative. Second, cultural distance is divided into organizational and national cultures. Organizational culture includes the use of methodologies and project management practices. And, National distance is the group's norms, values and spoken languages. The approaches to alleviate cultural distance is designed in graphs to select the minimum cultural distance. Third, reducing distance by using synchronous communication, such as email and teleconferencing.

Mullick, Neel et al. [31] designed an experiment that holds teams from different countries. The teams practiced the methods of work distribution. The work was split into coherent packages of work to be done by teams. The goal was to satisfy resources needs for tasks development. The task allocation factors they considered are from literature. The method they used is distributing the work manually among teams to be achieved by needed skills.

Grinter et al. [32] presented an approach to deal with coordination issues in a distributed work environment.

Prikladnicki et al. [33] conducted their model as it evaluates different sites for defined allocation criteria. As a result, the cost and benefit analysis for every site is counted. However, dependencies between tasks are not considered.

2.2.5 Systematic Solutions

Lamersdorf et al. [11] proposed a Bayesian Network model that takes three sites and considers cost, time and quality factors. The model suggests the appropriate site for each task using probabilistic values.

Fernandez et al. [34] proposed a method that is semi-automatic to distribute tasks to location. They used genetic algorithms. The study did not show any demonstration or validation but only the idea.

2.2.6 Comparison & Conclusion

We notice that some studies considers multiple goals, while others do not. Also, some models do not include dependencies between tasks. In addition, some decision techniques are used manually and not systematically. In TABLE I, a comparison of some approaches that could be comparable for handling the criteria: multiple objectives supportive, characteristics of sites, dependencies between tasks, empirical investigation and machine learning involvement.

Table 2.1: Comparable Existing Approaches for Work Distribution
Abbreviations: (-) Not, (o) partly, (+) mostly, and (++) totally.

Existing Model	*Considers multiple goals*	*Characteristics of Tasks*	*Characteristics of Site*	*Dependencies between Tasks*	*Dependencies between Sites*	*Empirically validated*	*Machine Learning Based*
Mockus and Weiss [18]	-	O	O	-	+	O	-
Mullick, N. et al. [31]	-	+	+	+	-	+	-
Madachy, R. et al [29]	-	+	++	-	-	O	-
Setamanit, S et al. [25]	+	-	+	O	+	O	-
Prikladnicki, R. et al. [33]	+	+	+	-	+	O	-
Fernandez al. [34]	O	+	-	O	-	-	+

CHAPTER 3

PROPOSED METHODOLOGY

Global Software Development needs a task allocation model in which individual work distribution units are assigned to suitable geographically distributed sites. In this chapter, we present a new task allocation model for GSD. The task allocation model, as shown in Figure 3.1, consisites of three phases: namely, the Data Generation, the Model Training & Site Section Processes, and Task Assignment.

- **First,** In phase 1, GSD decision factors are identified as feature of the dataset. In this phase, the data generation process for those decision factors is described.

- **Second,** In phase 2, the dataset is used for training the prediction models, the ANN and KNN classification [35, 36]. Then, the project manager provides the test site data, and the model predicts the fit sites for GSD.

- **Finally,** In phase 3, the task allocation model uses Simplex algorithm [37] to assign project tasks to the fit sites predicted in phase 2.

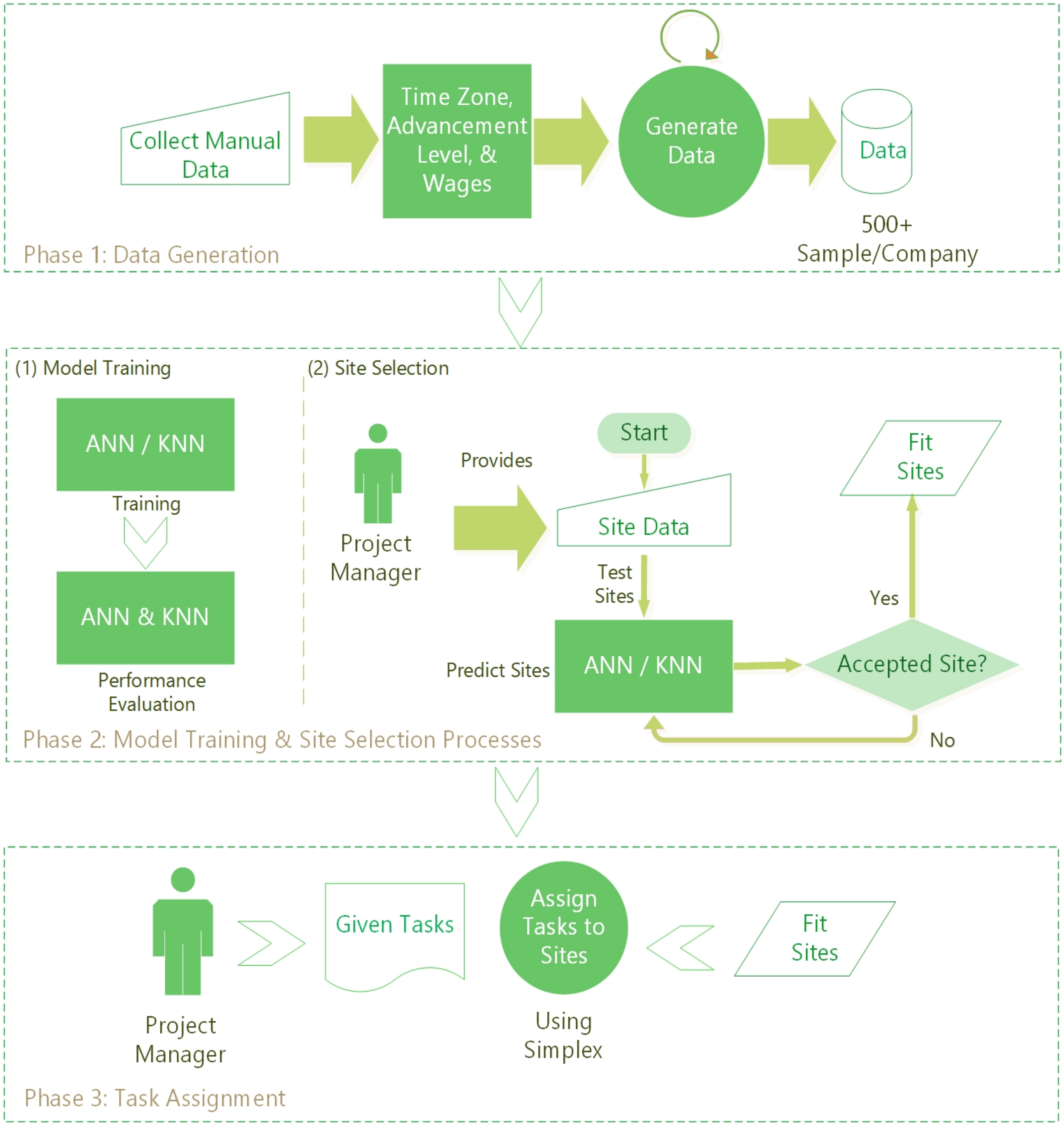

Figure 3.1: The Proposed Model General Phases

In the next sections, more details are for all phases of the model.

3.1 Phase 1: Data Generation

Researchers [19, 10, 28] have identified time difference, cost, skills, and quality as key task allocation factors in GSD projects. For example, Lamersdorf et al. empirically identified that time difference, quality and cost are the most important criteria for work distribution during GSD projects. Mahmood et al. [38] identified CMMI, time difference, and cost as key factors. In this theses, we only consider cost, CMMI and time difference as factors that potentially influence site selection process, as shown in Table 3.1. And the full generated dataset is included in Appendix A.

No.	Decision Factor	Description
1	CMMI	Capability Maturity Model Integration
2	Time differences	Time Zone Difference with respect to home site
3	Cost	The estimated cost for a site

Table 3.1: GSD Task Allocation Factors

A brief description about the those factors is as follows:

1. **CMMI** ranges from 1-5. It has five levels [39], initial level 1, Managed level 2, Defined level 3, Quantitatively Manage level 4, and Optimizing level 5. The higher the CMMI, the better the processes the company has.

2. **Time Differences** the time difference calculated between the home and remote site. The higher the time difference indicates more work shifts and productivity.

3. **Cost** is the cost estimate for a particular site indicated by the wages at the site's country.

However, to generate the data for the decision factors, some initial data are collected manually in order to use them in the systematic dataset generation. The initial data are shown in Table 3.2.

No.	Data Factor	Description
1	Country Population	The number of people in a country
2	Country Advancement Level	Developed, Developing & under developing countries
3	Time Zone	Minimum and maximum time zone for a country
4	Min. Wages	The minimum wage per country

Table 3.2: Data Generation Factors

The above list of characteristics are collected manually for a number of counties. Each country will have several sites in the dataset. The above characteristics will help us generate data for decision factors at Table 3.1.

3.1.1 Why Data Generation?

There is a lack of industrial data publicaly available for GSD projects. In this work, we systematically collect some initial data manually and use that to systematically generate the dataset for the machine learning algorithms. This is common in GSD as other researches as other researchers, e.i Lamersdorf [11], have used hypothetical data for their work.

3.1.2 Initial Data Factors

In this section, we present our method to generate the data. First, Some initial data factors are collected as shown at table 3.2. Second, the data set is generated using the initial data factors. The data collection process is described below.

Country Population

The population is obtained for each country [40] in the dataset. (We consider that the maximum population is 200 million for a countries having population more than that number).

Time Zone

The time zone is found for the country. If the country has cites with different time zone, the maximum and minimum time zone is found for that country.
The range between max and min time zone is used in time zone generation per site as discussed in section 3.1.3.

Country Advancement Level

One characteristic of the country is the advancement level. Each country is classified into one of three advancement levels, i.e *developed countries, developing countries and least developing countries.*

The Min. Wage

The Min. Wage for each country is obtained [41] (There is almost no information available about Max. Wages. However, the wages per country help us to estimate the cost at the remote site).

3.1.3 Data Generation Results

In this sub-setion, we present the data generation results. We generated more than 500 company record. Figure 3.2 describes the data generation process, where time zone, advancement level, and wages are collected. After that, they are used to generate the dataset.

Figure 3.2: Data Generation Process

The description of the dataset is shown in Table 3.3.

Data Set Description	Number
Countries	30
Companies	506
Recommended Companies	239
Non-Recommended Companies	267

Table 3.3: Data Set Description

The description of how the data factors are generated is as follows:

Number of Companies per country

For each country, there is a number of companies. Countries have different number of companies based on the **population of the country and the advancement level**.

Number of Companies $= \sum_{k=1}^{n}(a\% * P * b\% * AL * R(c,d))$

- $\boldsymbol{n}$ is number of countries.
- **P** is population, **AL** is AdvancementLevel, and **R** is a random number.
- **a,b,c,d** are constants obtained experimentally to produce realistic data.

Table 5.4 and Table 5.5 to describes two examples.

Country	Germany
Population	80 million
Developed Level	3
Number of Companie	28

Table 3.4: Example 1 for Number of Companies per Country

Country	Sudan
Population	37 million
Developed Level	1
Number of Companies	2

Table 3.5: Example 2 for Number of Companies per Country

The Cost

The cost is an estimate of the cost at one site. The scale is in dollar. The cost factor is based on the **advancement level and the country Min. Wage** as follows:

Cost $= \sum_{k=1}^{n}(AL * (C * Mw) * R)$

- $\boldsymbol{n}$ is number of companies in a country.
- **AL** is the AdvancementLevel.
- **C** is constant because $Min.Wage$ is the least paid wage.
- **MW** is the Minimum wage.
- **R** is a random number between 0 and 1. that provides variety in wages among different sites in one country.

The CMMI

The Capability Maturity Model (CMMI) is a process improvement training required in software development contracts. It is claimed to be used to guide process improvement across a project, division, or an entire organization. [39]. We use the CMMI as a representation of the quality that the client company will provide for us and so we do not use the term *quality* because it is quite general.

The CMMI level ranges from 1 to 5. In our work, we assume that a site from a higher advancement level country has a higher CMMI level as shown at table 3.6.

CMMI Level	Least-Developing countries	Developing Countries	Developed Countries
CMMI 1	30%	5%	5%
CMMI 2	30%	30%	5%
CMMI 3	30%	30%	50%
CMMI 4	5%	30%	20%
CMMI 5	5%	5%	20%

Table 3.6: CMMI distribution for countries

Time Zone Difference

We know that some big countries have cities with different time zones. So, the time zone maximum and minimum for such those countries is collected. Now, to estimate the time zone for a given city in that country, a random number is generated in between minimum and maximum time zone and it is considered as the time zone for that city.

In example See at Table 3.7, the maximum time zone for Australia is UTC+11, and the minimum time zone is UTC+5. The generated time zone for a given city in Australia is UTC+6 which is in between UTC+5 and UTC+11.

The same idea is presented for Canada country at Table 3.8.

Country	Australia
Max time zone	UTC+11
Min time zone	UTC+5
Generated time zone for a site	UTC+6

Table 3.7: Example 1 of time zone generation for a site

Country	Canada
Max time zone	UTC-3
Min time zone	UTC-8
Generated time zone for a site	UTC-5

Table 3.8: Example 2 of time zone generation for a site

Now, the time difference with respect with Home site is calculated (in this research, UTC+3 is considered as the time zone for our home site). In example at Table 3.9, The time zone difference between the site (MB1 company) and our site is 9 hours.

Country	Canada
Company Name	MB1
Generated Time Zone	UTC-6
Time Difference with Home Site	9 hour

Table 3.9: Example 1 of time zone difference calculation for a site

Country	Brazil
Company Name	VU9
Generated Time Zone	UTC-4
Time Difference with Home Site	7 hour

Table 3.10: Example 2 of time zone difference calculation for a site

The Data Class and Preprocessing

The sites at the dataset are classified as either 0 or 1. That is, the site achieving the average value of all training factors is classified as class 1, otherwise as class 0. The class 1 means that the site is recommended whereas class 0 means that the site is not recommended.

Now, the class is added to all generated companies records. After that, we added some noise to the data to avoid perfect learning because the class decision is clear. The purpose of the noise is to simulate real data. Figure 3.3 displays the ratio of noise on original data.

The added noise is the normal distribution noise (Gaussian Noise [42]). The reason behind selecting this type of noise is that the Cost and the CMMI factors are already assumed to be values that are normally distributed. That is, most of sites have CMMI around 2-4 level out of 1-5 range in level. Also, most sites are assumed to be asking for average price. However, for time-zone difference, no noise is added because its values are derived from real time zones.

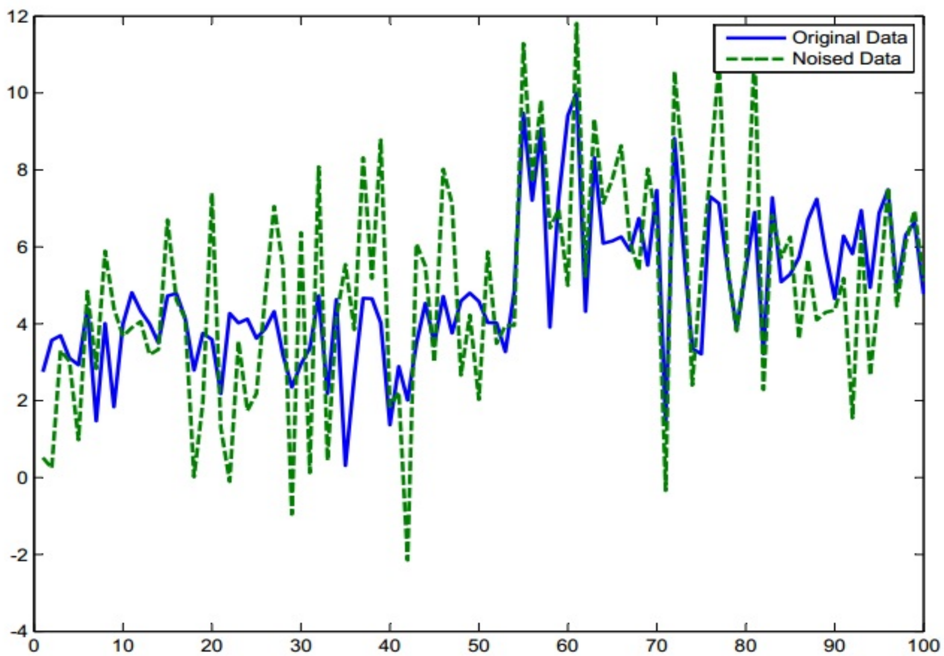

Figure 3.3: Gaussian Noise added to the Data

3.2 Phase 2: Model Training & Site Selection Process

As part of this work, the task allocation model first predicts the fit sites. To achieve that, first the model is trained and the training process is presented in this section. After that, the model becomes ready to predict fit sites.

The classification algorithms used to predict the fit sites are the ANN and KNN classification [43, 36]. We also do evaluation and compare performance of both machine learning techniques as described at Figure 3.4.

Figure 3.4: The Site Selection Workflow

The amount of data used for training the model is 80% of the dataset, whereas 20% of the data is used for testing. The cross-validation is used. The best model will be decided based on the high accuracy.

3.2.1 Artificial Neural Networks

The first prediction we include in our model is the Artificial Neural Networks (ANN). It is a machine learning technique that has learning abilities, and it is helpful in solving problems with uncertain conditions. ANN is constructed from a neuron processing elements. The elements are connected by a network of connections where each connection is assigned a weight. Furthermore, ANN is used as a classifier. A classifier propagates input vectors to output vectors. The network learns the input along with their output values. Once the network has learned, it can be used as a predictor for future data. [43]. Software engineers have taken the advantage of ANN to solve some problems such as defect prediction [44] and development effort estimation [35].

We use the data set that we described in 3.1 section as an input for the neural networks model.

3.2.2 K-Nearest Neighbors Classification

The second prediction model we include in our model is the K-Nearest Neighbors Classification. It is a classification algorithm where it classifies data into some classes. For each example in the data set, the k closest members are specified. A Euclidean Distance measure is used to calculate how close each member of the training set is to the target row that is being examined [36]. The data set that we generate will be used as an input to this classification algorithm. This algorithm will output two classes that is either recommended site or not to assign a breakdown task to.

3.2.3 Results and Evaluation

In this section, the accuracy of the techniques is investigated, that considers ratio of true classified over misclassified data. MATLAB R2012b is used to simulate the proposed task allocation decision model.

A feed-forward ANN with 3 inputs, one hidden layer and 1 output layer is used. The sigmoid function is used as an activation function for the ANN. The ANN parameters and the number of neurons are optimized. As a controlled experiment, at each time there is only one parameter is variable whereas other parameters are fixed. Figure 3.5 to shows the optimum number of neurons for the hidden layer of the ANN. The graph shows that the number of neurons approaches 13, the model gives the highest accuracy 96.91% and minimized error rate 3.09%. While increasing the number of neurons over than 13, the neural network starts over-fitting being producing poor predictive performance. So for this problem that we solve, 13 hidden neurons maximizes the ANN prediction accuracy.

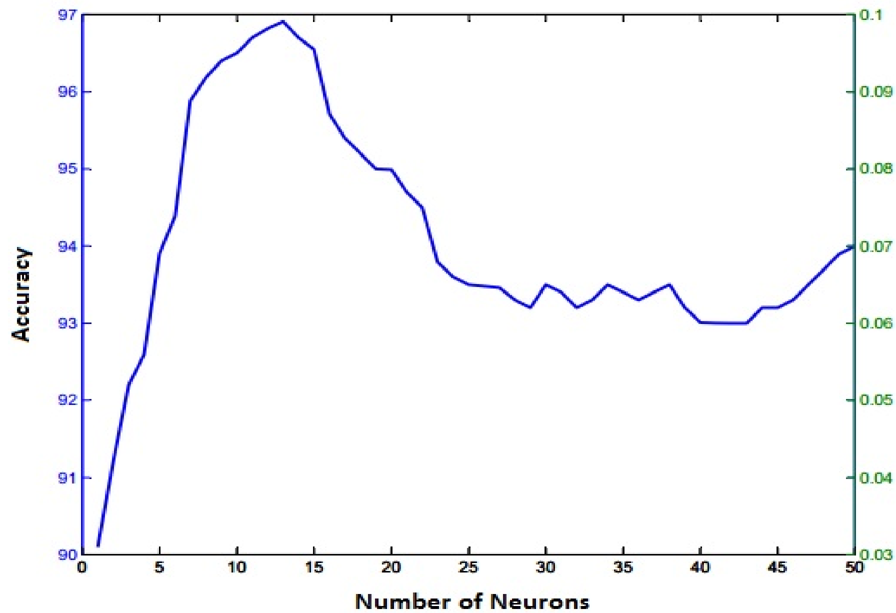

Figure 3.5: Number of Neurons Variations and Accuracy for ANN

For the KNN classification, two parameters are variable, the distance and k. So that first, the k is set variable and the distance fixed. Second, the distance is set variable and k is fixed. Figure 3.6 illustrates the progress of accuracy so that the KNN classifier produces the best accuracy at k =2.

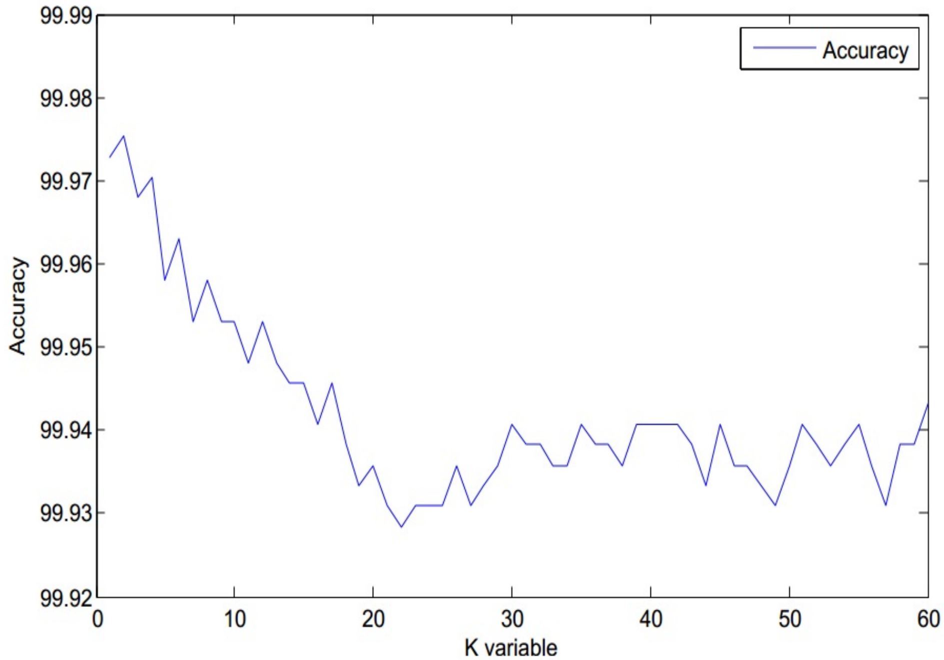

Figure 3.6: Monitoring KNN performance while K is variable

On the other hand, distances induced in this study are Euclidean, City block, Chebychev and MinKowski. They almost produce the similar accuracies while Euclidean distance produces quite higher accuracy among all other distances as shown at table 3.12.

KNN Distance	Error Rate	Accuracy
Euclidean	1.145%	98.855%
Cityblock	1.151%	98.849%
Chebychev	1.161%	98.839%
Minkowski	1.147%	98.853%

Table 3.11: KNN Distance Optimization

The comparison between ANN and KNN, shown at Table 3.13, presents that

KNN outperformed the ANN with difference around 1.945% in accuracy.

Algorithm	Error Rate	Accuracy
ANN	3.09%	96.91%
KNN	0.145%	98.855%

Table 3.12: ANN and KNN Performance Comparison

3.2.4 Site Selection Process

In the previous section, the model training and accuracy measurement were presented. Now, the decision model is ready to be used for the site selection. In the site selection process as descried at Fig. 3.6, the project manager provides the sites. This process predicts the fit-sites.

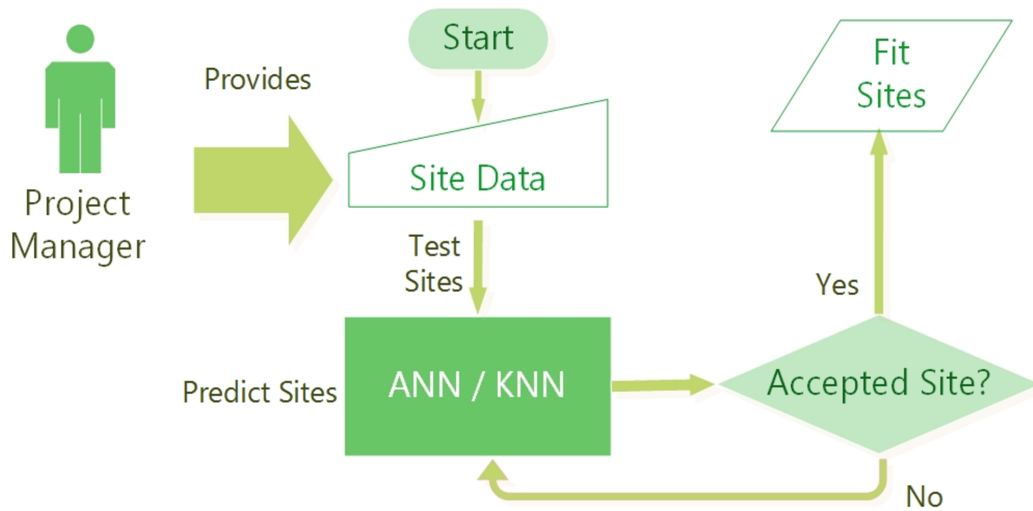

Figure 3.7: The Site Selection Workflow

Example of the site data is presented at table 3.13, where the data consists of three decision factors as features, which are the CMMI, the cost, the time zone difference and finally the site class (0 or 1). Each row in the data is specific information example about one company. For more details about the data, check section 3.1.3.

Site	**CMMI (1-5)**	**Cost(1-10)**	**Time Differene** (wrt Home Site)	**Class**
FH4	5	6	6	0
QT8	5	7	5	0
ZJ8	3	2.90	1	1
HZ5	2	3.01	5	1
RR8	4	2.62	2	1

Table 3.13: Introducing Company Information

An example of some data at Table 3.14, given by project manager to the model to make a decision on them. The data is classified and explanations for that classification is presented below:

- The two first companies in the table are classified as not recommended sites. One explanation is that they have too high values of cost.
- Company number 3 in the table is classified as recommended and that might be because of good indicators of CMMI and Cost.
- Company number 4, it is recommended based on its preferred values on cost and time zone difference.
- Whereas for company number 5, it is recommended based on its excellence in CMMI and cost although it has small time zone difference. However, this decision is based on the learning knowledge and similar data that the model has already seen before.

Site	**CMMI (1-5)**	**Cost(1-10)**	**Time Differene** (wrt Home Site)	**Class**
FH4	5	6.02	6	0
QT8	5	7.30	5	0
ZJ8	3	2.90	1	1
HZ5	2	3.01	5	1
RR8	4	2.62	2	1

Table 3.14: Deciding on companies

3.3 Phase 3: Task Assignment

Software development projects usually include several tasks. In this section, the method used to assign the project tasks automatically is presented. The output of the previous phase, *the fit sites*, is considered as one input in this phase, the task assignment process as shown at Figure 3.7. The second input is the tasks by the project manager. The method used is called Simplex algorithm which is discussed below.

Figure 3.8: Task Assignment to fit sites

3.3.1 Simplex Method

The Simplex Algorithm is a geometric algorithm for linear programming problems [37]. The algorithm implements constraints to maximize profit or immunize cost.

However, in this research, we use the simplex aglorithm to solve the task assignment as a special case of the transportation problem. The constrain is that each task should be assigned to one site. The goal is to provide us with a systematic task assignment while minimizing the total cost.

3.3.2 Task Assignment Scenarios

The input to the algorithm are three input variables: the sites, the tasks and the cost coefficients as follows:

- First, the fit-sites which are predicted by the prediction model at previous section 3.2.4 (The Site Selection Process).
- Second, the given tasks by the project manager.
- Finally, the man-hour quotes representing the cost coefficients.

The Excel Solver plug-in is used to implement the concept of simplex algorithm in this research. The method is represented in a table, where rows are tasks and columns are sites. The cells contain the man-hour cost. Capacity constraint is the constraint for the sites. The method output is solving the problem by assigning each task to one site while minimizing the total man-hour quotes. Below are two simplex examples.

Scenario 1

In this scenario, there are 4 sites, and 4 tasks. The man-hour quotes are in the table as shown at Table 3.15.

Tasks , Sites	Site 1	Site 2	Site 3	Site 4	Total
Task 1	69	63	41	40	1 site
Task 2	110	125	120	130	1 site
Task 3	98	86	94	88	1 site
Task 4	165	143	140	177	1 site
Tasks Assigned	1 task	1 task	1 task	1 task	4 tasks
Cost Assigned	110 hrs	86 hrs	140 hrs	40 hrs	376 hrs

Table 3.15: Tasks Assignment Scenario 1.

The method results at Table 3.15 displays that:

- Site 1 is assigned Task 2.
- Site 2 is assigned Task 3.
- Site 3 is assigned Task 4.
- Site 4 is assigned Task 1.

The minimum cost of assigning all tasks is 376 man-hour.

Scenario 2

In this second scenario, there are 4 sites, and 6 tasks. The man-hour quotes are in the table cells as shown at Table 3.16 This scenario demonstrates more solutions:

- The algorithm, in this scenario, does task assignment while the tasks are more than sites. It does not leave any tasks unassigned.

- The algorithm, in this scenario, includes constraints. For example, site 2, Site X, can to take up to 2 tasks only.

Sites →	**FH4 site**	**Site 2**	**Site 3**	**Site 4**	**Total**
Constraint →	**1 task**	**2 task**	**5 task**	**1 task**	**9 task**
Task 1	77	85	95	70	1 site
Task 2	120	105	114	102	1 site
Task 3	102	116	111	105	1 site
Task 4	28	27	31	37	1 site
Task 5	97	108	100	91	1 site
Task 6	87	74	90	78	1 site
Tasks Assigned	**1 task**	**2 tasks**	**2 tasks**	**1 task**	**6 tasks**
Cost Assigned	**102 hrs**	**179 hrs**	**131 hrs**	**70 hrs**	**482 hrs**

Table 3.16: Tasks Assignment

The method results at Table 3.16 displays that:

- Site 1 is assigned Task 3.
- Site 2 is assigned Task 2 & 6.
- Site 3 is assigned Task 4 & 5.
- Site 4 is assigned Task 1.

The minimum cost of assigning all tasks is 482 man-hour.

CHAPTER 4

THE CASE STUDY

We discussed the prediction models and the task assignment algorithm in the previous chapter. In this chapter, we present a real case study that is conducted in the decision model.

4.1 Project Manager Requirements

The case study is at the Architectural department at KFUPM. They want to develop an application that fulfills some requirements at the department. They want to use the GSD concept and know the best-fit sites to assign the development tasks to them. Furthermore, they need the model to predict fit sites and assign tasks to them.

4.2 The Case Study Workflow

The workflow of this case study is described at Figure 4.1. The process passes through several steps as following:

- First, we recieved the sites and tasks from the project manager along with the man-hours quotes.
- Second, the optimum sites are predicted, and they are consumed by the simplex algorithm as input along with the tasks.
- Lastly, the Simplex algorithm assigns the tasks to the sites.

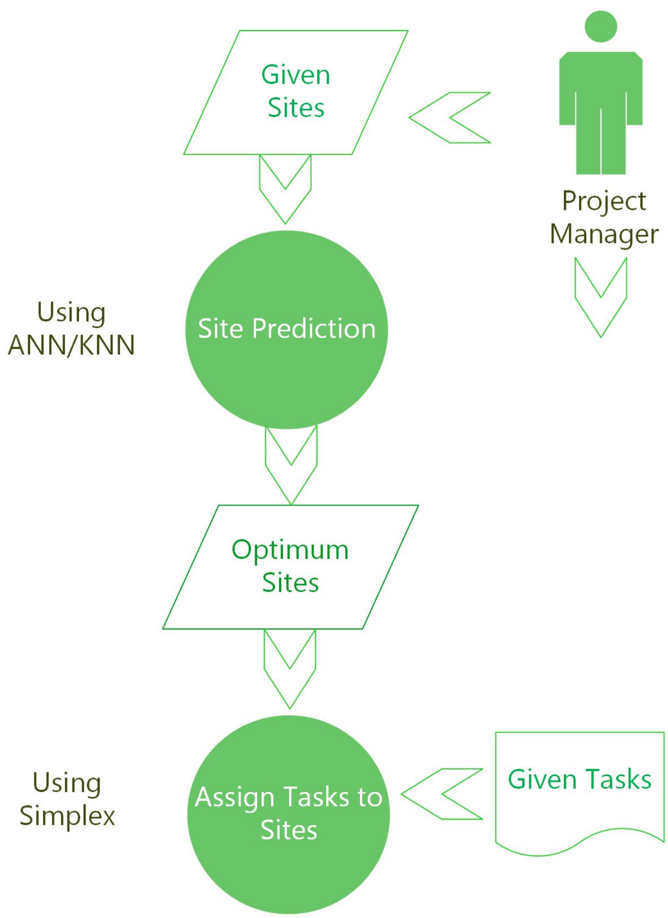

Figure 4.1: Workflow for Site Prediction and Tasks Assignment

4.3 The Sites

The description of the given sites is on table 4.1. It shows sites data: the *CMMI level*, the *cost scheme* and the *time zone differences*.

In this case study, there are 7 sites on the table. Theses sites will be optimized by our model to produce the fit sites.

No.	**Company**	**CMMI**	**Cost**	**Time Differene** (wrt to home site)
1	Poland	2	4.91	1
2	China	5	3.01	2
3	United States	4	5.52	6
4	Canada	4	2.19	5
5	China	5	7.39	5
6	United Kingdom	1	5.32	3
7	India	4	4.57	3

Table 4.1: Given Sites Data

g

4.4 The Tasks

The given tasks are 30 tasks as shown at Table 4.2. We assume that these tasks are independent.

No.	Task Name	China	U.S	Canada	India
1	Source node	69	63	41	60
2	User profile	140	125	135	130
3	Architect profile	98	125	94	98
4	Import SimaPro style data files	165	162	154	177
5	Import excel style data files	39	35	47	50
6	Import openLCA style data files	94	84	83	75
7	File type conversions	104	100	118	118
8	Create life cycle assessment flows	105	86	111	104
9	Define life cycle assessment processes	36	25	27	33
10	Complete life cycle assessment work flows	105	97	109	93
11	LCA screen layouts	94	85	95	87
12	LCA database implementation	49	53	53	52
13	LCA business logic and database integration	70	60	53	58

14	Import LCA standard methods	159	183	183	179
15	Sequential inventory calculation feature	145	129	114	137
16	Uncertainty calculation feature	129	139	124	136
17	Characterization feature	145	133	148	151
18	Environmental damage assessment feature	241	180	240	232
19	Normalization feature	191	181	194	209
20	Energy modeling assessment feature	162	168	168	181
21	Group analysis	299	297	276	295
22	Standard analysis	228	243	269	236
23	Graphics based analysis	264	231	215	273
24	Spread sheet based analysis	329	327	314	307
25	Export SimaPro analysis form	358	360	368	354
26	Export openLCA analysis form	372	394	374	399
27	Export excel analysis form	335	315	343	344

28	Export text analysis form	306	307	308	310
29	LCA forms integration	316	323	319	304
30	Important construction products	309	309	304	310

Table 4.2: Given Tasks

4.5 Case Study Results

In this sub-section, Table 4.3 displays the optimum sites. It is 4 out of 7 given sites that are recommended by our model. They are China, United States, Canada and India.

No.	**Company**	**CMMI**	**Cost**	**Time Differene** (wrt to home site)
1	China	5	3.01	2
2	United States	4	5.52	6
3	Canada	4	2.19	5
4	India	4	4.57	3

Table 4.3: Predicted Optimum Sites

Nex, the Simplex Algorithm uses those optimum sites and the task. Table 4.4 presents the simplex algorithm, where the columns are sites, rows are tasks, and the table's cell values are the man-hours required for each task. The algorithm assigns the tasks to sites while minimizing the man-hours total cost.

Site Name →	**China**	**U.S.**	**Canada**	**India**	**Total**
Constraint →	**5 task**	**6 task**	**4 tasks**	**15 task**	**30**
Task 1: Source node	69	63	41	60	1 site
Task 2: User profile	140	125	135	130	1 site
Task 3: Architect profile	98	125	94	98	1 site
Task 4: Import SimaPro style data files	165	162	154	177	1 site
Task 5: Import excel style data files	39	35	47	50	1 site
Task 6: Import openLCA style data files	94	84	83	75	1 site
Task 7: File type conversions	104	100	118	118	1 site
Task 8: Create life cycle assessment flows	105	86	111	104	1 site
Task 9: Define life cycle assessment processes	36	25	27	33	1 site
Task 10: Complete life cycle assessment work flows	105	97	109	93	1 site
Task 11: LCA screen layouts	94	85	95	87	1 site
Task 12: LCA database implementation	49	53	53	52	1 site
Task 13: LCA business logic and database integration	70	60	53	58	1 site

Task 14: Import LCA standard methods	159	183	183	179	1 site
Task 15: Sequential inventory calculation feature	145	129	114	137	1 site
Task 16: Uncertainty calculation feature	129	139	124	136	1 site
Task 17: Characterization feature	145	133	148	151	1 site
Task 18: Environmental damage assessment feature	241	180	240	232	1 site
Task 19: Normalization feature	191	181	194	209	1 site
Task 20: Energy modeling assessment feature	162	168	168	181	1 site
Task 21: Group analysis	299	297	276	295	1 site
Task 22: Standard analysis	228	243	269	236	1 site
Task 23: Graphics based analysis	264	231	215	273	1 site
Task 24: Spread sheet based analysis	329	327	314	307	1 site
Task 25: Export SimaPro analysis form	358	360	368	354	1 site
Task 26: Export openLCA analysis form	372	394	374	399	1 site

Task 27 : Export excel analysis form	335	315	343	344	1 site
Task 28 : Export text analysis form	306	307	308	310	1 site
Task 29 : LCA forms integration	316	323	319	304	1 site
Task 30 : Important construction products	309	309	304	310	1 site

Table 4.4: Given Tasks

The same Table 4.4 displays the results of the assignment. In that table, the values that are highlighted represent the assignment of a task to a site. For example, China is assigned 6 sites. United States is assigned 12 tasks. Canada is assigned 7 tasks and India is assigned 5 tasks. The minimum assignment cost in terms of man-hours, is 5124 hour.

CHAPTER 5

DISCUSSION

In this section, some additional characteristics of the methodology is outlined that make it more attractive for practical use compared with the current techniques for making task allocation decisions.

- **Multi-Factor Support:** Unlike some existing methods that consider only single purpose. Our model considers multiple factors where most existing models focus on cost or dependencies only [28, 27, 19]. Moreover, Setamanit, et al.'s model [25] considered task dependencies only. Also, Mockus and Weiss [18], is allocating work items based on their modularity. Madachy, Ray [29] based his model on the cost for work distribution. Jabangwe, R. and Smite, D. [30] discussed different stratifies that affect the eventual product quality as that considering the quality factor for decision. Other factors are not fulfilled.

  Our approach considers multiple factors Cost, CMMI level, and Time Zone differences.

- **Systematic Decision:** Some existing models are not making task allocation decisions systematically. Beecham et al. [16]'s method gives general statement or recommendations to project managers with an aim to facilitate appropriate site selection that best meets business needs of the project. Wickramaarachchi and Lai [10] identify work-specific characteristics, relation between the work phases, dependency between sites and the site-specific characteristics. Decision is made by putting all into one formula and to that for each site manually.

  Our model uses well-established machine learning techniques, with enough accuracy in predictions that provides a systematic decision for the project manager.

- **Scalability:**

  The techniques used in this approach for site selection decision is based on machine learning, where it can make prediction based on large number of factors [45]. That is important because factors for making task allocation are increasing [19]. Example of those factors are: Collaboration Maturity, Reliability, Expertise, Proximity to client, Personal trust/contacts, Proximity to market, Availability, Development quality, Preparation at site and Political Decisions.

- **Criteria Definition:** There should be more research, surveys and interviews with professionals that are focusing on enhancing the decision factors definition.

- **Single Scale:** Task assignment, based on Simplex algorithm. The algorithm decides on either hours or dollar scales, but not together. Other algorithms need to be investigated to deal with multiple scales together.

- **Special Cases:** This is general case where we optimize the sites based on all decision factors. A project manager might still need to optimize three factors but giving priority to one or tow of them which is not achieved in our case.

CHAPTER 6

CONCLUSION, FUTURE WORK & APPENDICES

6.1 Conclusion

Today, more and more companies are using the concept of global software development. As a result of the technology advancement we observe today, GSD became more popular because of the improvement of Internet communication and software development technologies. In the other hand, companies that provide global software development services are increasing also around the world.

Some of benefits of adopting GSD is accessing highly qualified people around the world, working almost 24/7 or achieving follow-the-sun concept and looking for lower development wages at some countries. In this search, a task allocation decision model in global software development is introduced. This research considers those decision factors and introduces a novel idea of predicting suitable companies to assign tasks to. The mode is based using machine learning techniques, the ANN and KNN. Both techniques are trained and tested on the data and the results stated that the KNN outperformed the ANN. This provides the flexibility for professionals to select the appropriate algorithm for specific problems. The model is also supported by a systematic task assignment part which allow tasks to be assigned automatically to sites.

Actually collecting real data needs huge efforts and travel to reach global sites. However, the model will need more R&D to collect real data and apply them to it.

6.2 Future Work

- In future, we plan to validate our model on various data sets.
- Our approach is conducted on systematically generated data. We plan to collect real data from industry.
- The task dependency is and additional solution that we plan to research more on it, and adopt it to our model.
- Do more research on how can we allow the model to give priority to some decision factors based on project manager requirements.
- The model helps professionals to decide on multiple factors. But still we plan to produce other versions of the model to deal with single factors based on project managers requirements.

- Task assignment considers site capacity constraint. In future, we plan to express more constraints by project manager.

6.3 REFERENCES

REFERENCES

[1] Christof Ebert. *Global Software and IT: A guide to Distributed Development, Projects and Outsourcing*. IEEE Computer Society, 2012.

[2] T. Kern and L. Willcocks. Exploring information technology outsourcing relationships: theory and practice. *Journal of Strategic Information Systems*, 9:321 – 350, 2000.

[3] A.A. Bush, A. Tiwana, and H. Tsuji. An empirical investigation of the drivers of software outsourcing decisions in japanese organizations. *Information and Software Technology Journal*, 50(6):499–510, 2008.

[4] M. Niazi, S. Mahmood, M. Alshayeb, M. Rehan Riaz, K. Faisal, and N. Cerpa. Challenges of project management in global software development: Initial results. In *Science and Information Conference (SAI), 2013*, pages 202–206, Oct 2013.

[5] L McLaughlin. An eye on india: Outsourcing debate continues. *IEEE Software*, 20(3):114–117, 2003.

[6] S. Islam, M.M.A. Joarder, and S.H. Houmb. Goal and risk factors in offshore

outsourced software development from vendor's viewpoint. In *Global Software Engineering, 2009. ICGSE 2009. Fourth IEEE International Conference on*, pages 347–352, July 2009.

[7] M. Cataldo, M. Bass, J.D. Herbsleb, and L. Bass. On coordination mechanisms in global software development. In *Global Software Engineering, 2007. ICGSE 2007. Second IEEE International Conference on*, pages 71–80, Aug 2007.

[8] Mary C. Lacity and Joseph W. Rottman. Effects of offshore outsourcing of information technology work on client project management. *Strategic Outsourcing: An International Journal*, 2(1):4–26, 2008.

[9] J. Stark, M. Arlt, and D.H.T. Walker. Outsourcing decisions and models - some practical considerations for large organizations. In *Global Software Engineering, 2006. ICGSE '06. International Conference on*, pages 12–17, Oct 2006.

[10] D. Wickramaarachchi and R. Lai. A method for work distribution in global software development. In *Advance Computing Conference (IACC), 2013 IEEE 3rd International*, pages 1443–1448, 2013.

[11] Ansgar Lamersdorf, Jürgen Münch, and Dieter Rombach. A decision model for supporting task allocation processes in global software development. In *Product-Focused Software Process Improvement*, pages 332–346. Springer, 2009.

[12] Ravi Aron and Jitendra V Singh. Getting offshoring right. *Harvard business review*, 83(12):135, 2005.

[13] Najmul Huda, Nazmun Nahar, Jaak Tepandi, and P Singha Deo. Key barriers for global software product development organizations. In *Management of Engineering & Technology, 2009. PICMET 2009. Portland International Conference on*, pages 1081–1087. IEEE, 2009.

[14] Pär J Ågerfalk, Brian Fitzgerald, Helena Holmström Olsson, and Eoin Ó Conchúir. Benefits of global software development: the known and unknown. In *Making Globally Distributed Software Development a Success Story*, pages 1–9. Springer, 2008.

[15] Daniela Damian and Deependra Moitra. Guest editors' introduction: Global software development: How far have we come? *Software, IEEE*, 23(5):17–19, 2006.

[16] Sarah Beecham, John Noll, Ita Richardson, and Deepak Dhungana. A decision support system for global software development. In *Global Software Engineering Workshop (ICGSEW), 2011 Sixth IEEE International Conference on*, pages 48–53. IEEE, 2011.

[17] Erran Carmel and Ritu Agarwal. Tactical approaches for alleviating distance in global software development. *Software, IEEE*, 18(2):22–29, 2001.

[18] Audris Mockus and David M Weiss. Globalization by chunking: a quantitative approach. *Software, IEEE*, 18(2):30–37, 2001.

[19] Ansgar Lamersdorf, Jürgen Munch, and Dieter Rombach. A survey on the state of the practice in distributed software development: Criteria for task allocation. In *Global Software Engineering, 2009. ICGSE 2009. Fourth IEEE International Conference on*, pages 41–50. IEEE, 2009.

[20] James D. Herbsleb, Audris Mockus, Thomas A. Finholt, and Rebecca E. Grinter. Distance, dependencies, and delay in a global collaboration. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, CSCW '00, pages 319–328, New York, NY, USA, 2000. ACM.

[21] P. Zencke. Communication in software development. 2011.

[22] et al Grinter, R.E. The geography of coordination: Dealing with distance in r&d work. *Proceedings of GROUP99. ACM Press, New York, 1999, pp. 306315.*, 1999.

[23] J.D. Herbsleb and A. Mockus. An empirical study of speed and communication in globally distributed software development. *IEEE Transactions on Software Engineering, Vol. 29, no. 3, pp. 481494*, 2003.

[24] James M Hussey and Steven E Hall. *Managing Global Development Risk*. CRC Press, 2007.

[25] Siri-on Setamanit, Wayne Wakeland, and David Raffo. Using simulation to evaluate global software development task allocation strategies. *Software Process: Improvement and Practice*, 12(5):491–503, 2007.

[26] Supraja Doma, Larry Gottschalk, Tetsutaro Uehara, and Jigang Liu. Resource allocation optimization for gsd projects. In *Computational Science and Its Applications–ICCSA 2009*, pages 13–28. Springer, 2009.

[27] Matthew Bass and Daniel Paulish. Global software development process research at siemens. In *Third International Workshop on Global Software Development*, pages 8–11, 2004.

[28] Ansgar Lamersdorf, J Munch, and Dieter Rombach. Towards a multi-criteria development distribution model: An analysis of existing task distribution approaches. In *Global Software Engineering, 2008. ICGSE 2008. IEEE International Conference on*, pages 109–118. IEEE, 2008.

[29] Ray Madachy. Distributed global development parametric cost modeling. In *Software Process Dynamics and Agility*, pages 159–168. Springer, 2007.

[30] R. Jabangwe and D. Smite. Decision support for offshore insourcing software development. In *Global Software Engineering Workshop (ICGSEW), 2011 Sixth IEEE International Conference on*, pages 111–113, Aug 2011.

[31] Neel Mullick, Matthew Bass, Z Houda, Paulish Paulish, and Marcelo Cataldo. Siemens global studio project: experiences adopting an integrated gsd infrastructure. In *Global Software Engineering, 2006. ICGSE'06. International Conference on*, pages 203–212. IEEE, 2006.

[32] Rebecca E Grinter, James D Herbsleb, and Dewayne E Perry. The geography of coordination: dealing with distance in r&d work. In *Proceedings of the*

international ACM SIGGROUP conference on Supporting group work, pages 306–315. ACM, 1999.

[33] Rafael Prikladnicki, Jorge Luis Nicolas Audy, and Roberto Evaristo. A reference model for global software development: findings from a case study. In *Global Software Engineering, 2006. ICGSE'06. International Conference on*, pages 18–28. IEEE, 2006.

[34] J. Fernandez and M. Basavaraju. Task allocation model in globally distributed software projects using genetic algorithms. In *Global Software Engineering (ICGSE), 2012 IEEE Seventh International Conference on*, pages 181–181, Aug 2012.

[35] Heejun Park and Seung Baek. An empirical validation of a neural network model for software effort estimation. *Expert Systems with Applications*, 35(3):929–937, 2008.

[36] http://www.solver.com/k-nearest-neighbors-k-nn-classification-intro.

[37] Katta G Murty. *Linear programming*, volume 57. Wiley New York, 1983.

[38] Sajjad Mahmood, Sajid Anwer, Mahmood Niazi, Mohammad Alshayeb, and Ita Richardson. Identifying the factors that influence task allocation in global software development: Preliminary results. In *19th International Conference on Evaluation and Assessment in Software Engineering, EASE 2015*, Nanjing, China, 04/2015 2015.

[39] CMMI Product Team. Cmmi for development, version 1.2. 2006.

[40] U.s. and world population clock.

[41] Official exchange rate (lcu per us, $periodaverage$). 2013.

[42] Carl Edward Rasmussen. Gaussian processes for machine learning. 2006.

[43] James A Anderson and Joel Davis. *An introduction to neural networks*, volume 1. MIT Press, 1995.

[44] Venkata Udaya B Challagulla, Farokh B Bastani, I-Ling Yen, and Raymond A Paul. Empirical assessment of machine learning based software defect prediction techniques. *International Journal on Artificial Intelligence Tools*, 17(02):389–400, 2008.

[45] Tim Hill, Leorey Marquez, Marcus O'Connor, and William Remus. Artificial neural network models for forecasting and decision making. *International Journal of Forecasting*, 10(1):5 – 15, 1994.

6.4 Appendix A: The DataSet

This appendix contains the data collected for this study. The initial data are: Max time zone, Min time zone, population (in Millions), Min Wage and Advancement Level. The generated data are: The number of companies, the company names, the site time zone, and the company size. Those data are all involoved in generating the decision factor data in this table, the Cost, the CMMI, and the Time Zone Differences.

Initial Data													Training Data			
No.	Country Code	Country Name	MIN Time Zone	MAX Time Zone	Population (Million)	Min Wage	Advancement Level	No. of Companies	Company Name	Site	Site Time Zone	Company Size	Cost (1-10)	CMMI (1-5)	Time Differene wrt Home site (0-8)	Class
1			4	11	200	7.2	3		ON7	Site1	11	715	4.06	5	8	0
2			4	11	200	7.2	3		KP4	Site2	4	1979	3.29	5	1	0
3			4	11	200	7.2	3		EW7	Site3	11	1694	4.58	5	8	0
4			4	11	200	7.2	3		FC3	Site4	6	1705	4.37	5	3	0
5			4	11	200	7.2	3		CV2	Site5	4	1612	2.93	5	1	1
6			4	11	200	7.2	3		QM3	Site6	9	1688	2.98	5	6	1
7			4	11	200	7.2	3		KJ5	Site7	8	533	3.75	5	5	0
8			4	11	200	7.2	3		PK1	Site8	8	1575	3.29	5	5	0
9			4	11	200	7.2	3		DM6	Site9	8	647	2.68	5	5	1
10			4	11	200	7.2	3		RV1	Site10	10	779	2.39	5	7	1
11			4	11	200	7.2	3		ZN8	Site11	4	370	2.74	5	1	1
12			4	11	200	7.2	3		KE1	Site12	6	781	4.56	5	3	0
13			4	11	200	7.2	3		CR6	Site13	11	126	4.54	4	8	0
14			4	11	200	7.2	3		PU7	Site14	4	71	2.80	4	1	1
15			4	11	200	7.2	3		YT8	Site15	5	67	3.90	4	2	0
16			4	11	200	7.2	3		QE6	Site16	10	164	3.72	4	7	0
17			4	11	200	7.2	3		NI4	Site17	10	78	2.15	4	7	1
18			4	11	200	7.2	3		BG6	Site18	11	240	4.54	4	8	0
19			4	11	200	7.2	3		UU7	Site19	6	54	3.93	4	3	0
20			4	11	200	7.2	3		CE9	Site20	8	147	4.08	4	5	0
21			4	11	200	7.2	3		PH2	Site21	6	74	3.46	4	3	0
22			4	11	200	7.2	3		GW9	Site22	11	115	3.83	4	8	0
23			4	11	200	7.2	3		SG7	Site23	6	213	3.78	5	3	0
24			4	11	200	7.2	3		ET9	Site24	5	101	3.62	3	2	0
25			4	11	200	7.2	3		BX2	Site25	4	246	1.25	3	1	1
26			4	11	200	7.2	3		NP5	Site26	7	88	4.29	3	4	0

27	1	United States	4	11	200	7.2	3	54	HC2	Site27	10	132	3.59	3	7	0
28			4	11	200	7.2	3		WY3	Site28	10	173	3.82	3	7	0
29			4	11	200	7.2	3		NN1	Site29	4	215	3.63	3	1	0
30			4	11	200	7.2	3		SV1	Site30	5	217	2.22	3	2	1
31			4	11	200	7.2	3		TC6	Site31	11	134	3.55	3	8	0
32			4	11	200	7.2	3		TC6	Site32	6	133	4.43	3	3	0
33			4	11	200	7.2	3		AV6	Site33	7	127	3.19	3	4	0
34			4	11	200	7.2	3		LX7	Site34	7	146	3.11	3	4	0
35			4	11	200	7.2	3		SY3	Site35	11	180	3.85	3	8	0
36			4	11	200	7.2	3		MF2	Site36	4	118	4.27	3	1	0
37			4	11	200	7.2	3		EA9	Site37	10	120	3.97	3	7	0
38			4	11	200	7.2	3		XJ2	Site38	9	88	3.83	3	6	0
39			4	11	200	7.2	3		GF8	Site39	5	98	4.53	3	2	0
40			4	11	200	7.2	3		BL8	Site40	11	245	4.45	3	8	0
41			4	11	200	7.2	3		KG4	Site41	9	68	2.08	3	6	1
42			4	11	200	7.2	3		HY4	Site42	4	207	4.46	3	1	0
43			4	11	200	7.2	3		BG6	Site43	9	56	2.34	3	6	1
44			4	11	200	7.2	3		VV2	Site44	5	249	4.40	3	2	0
45			4	11	200	7.2	3		EF2	Site45	6	208	3.95	3	3	0
46			4	11	200	7.2	3		JJ4	Site46	10	143	4.46	3	7	0
47			4	11	200	7.2	3		ZU1	Site47	8	207	3.68	3	5	0
48			4	11	200	7.2	3		VA2	Site48	4	59	4.02	3	1	0
49			4	11	200	7.2	3		VN8	Site49	5	78	3.39	2	2	0
50			4	11	200	7.2	3		DS8	Site50	11	61	4.46	2	8	0
51			4	11	200	7.2	3		XG4	Site51	7	118	2.06	2	4	1
52			4	11	200	7.2	3		HV7	Site52	5	172	0.97	1	2	0
53			4	11	200	7.2	3		GL8	Site53	5	135	4.44	1	2	0
54			4	11	200	7.2	3		XM3	Site54	10	156	4.19	1	7	0
55			5	11	23	16	3		XJ8	Site1	6	875	7.06	1	3	0
56			5	11	23	16	3		FL3	Site2	10	925	5.88	2	7	0
57			5	11	23	16	3		MK9	Site3	8	1014	8.44	3	5	0
58	2	Australia	5	11	23	16	3	8.28	VA6	Site4	7	186	6.66	4	4	0
59			5	11	23	16	3		KD8	Site5	7	165	10.00	4	4	0
60			5	11	23	16	3		VM8	Site6	8	55	7.39	3	5	0

61			5	11	23	16	3		BS4	Site7	7	229	4.13	5	4	0
62	3	Switzerland	1	1	8	13	3	12	NO7	Site1	1	982	7.22	4	2	0
63			1	1	8	13	3		PQ2	Site2	1	1558	7.67	5	2	0
64			1	1	8	13	3		XA9	Site3	1	1799	8.20	4	2	0
65			1	1	8	13	3		GH1	Site4	1	73	7.25	5	2	0
66			1	1	8	13	3		YS2	Site5	1	216	6.53	3	2	0
67			1	1	8	13	3		NP2	Site6	1	159	4.41	3	2	0
68			1	1	8	13	3		QM6	Site7	1	249	5.90	3	2	0
69			1	1	8	13	3		YW8	Site8	1	177	8.08	3	2	0
70			1	1	8	13	3		GR3	Site9	1	58	4.59	3	2	0
71			1	1	8	13	3		GI2	Site10	1	88	6.75	3	2	0
72			1	1	8	13	3		FI4	Site11	1	160	6.83	1	2	0
73			1	1	8	13	3		DX4	Site12	1	242	4.44	2	2	0
74	4	Germany	1	2	80	11	3	28.8	WS7	Site1	1	377	3.71	5	2	0
75			1	2	80	11	3		UQ4	Site2	1	697	5.25	5	2	0
76			1	2	80	11	3		FZ3	Site3	1	1202	3.84	5	2	0
77			1	2	80	11	3		TB8	Site4	1	1321	4.96	5	2	0
78			1	2	80	11	3		OT4	Site5	2	1858	5.10	5	1	0
79			1	2	80	11	3		GM2	Site6	1	1012	6.47	4	2	0
80			1	2	80	11	3		RG9	Site7	2	530	4.10	4	1	0
81			1	2	80	11	3		JD3	Site8	1	1705	3.48	4	2	0
82			1	2	80	11	3		XO2	Site9	1	149	6.56	4	2	0
83			1	2	80	11	3		NC4	Site10	1	115	3.48	4	2	0
84			1	2	80	11	3		PV3	Site11	2	171	1.81	3	1	1
85			1	2	80	11	3		DD7	Site12	1	231	4.04	3	2	0
86			1	2	80	11	3		PX1	Site13	1	250	4.02	3	2	0
87			1	2	80	11	3		FY4	Site14	1	226	4.99	3	2	0
88			1	2	80	11	3		SS7	Site15	2	120	6.80	3	1	0
89			1	2	80	11	3		FO5	Site16	2	235	5.64	3	1	0
90			1	2	80	11	3		JU5	Site17	1	92	3.70	3	2	0
91			1	2	80	11	3		BA2	Site18	2	114	4.54	3	1	0
92			1	2	80	11	3		SE7	Site19	1	87	4.00	3	2	0
93			1	2	80	11	3		MF2	Site20	2	187	3.15	3	1	0
94			1	2	80	11	3		SE2	Site21	1	201	3.36	3	2	0
95			1	2	80	11	3		HN7	Site22	2	247	4.73	3	1	0

96			1	2	80	11	3		OK4	Site23	1	72	3.95	3	2	0
97			1	2	80	11	3		LQ3	Site24	2	128	6.60	3	1	0
98			1	2	80	11	3		ZM6	Site25	2	114	5.76	2	1	0
99			1	2	80	11	3		AJ2	Site26	2	98	4.83	2	1	0
100			1	2	80	11	3		UH8	Site27	2	205	1.48	1	1	0
101			1	2	80	11	3		NR5	Site28	2	54	6.37	1	1	0
102			1	2	80	11	3		FH6	Site29	1	131	4.98	2	2	0
103			1	2	80	11	3		QD7	Site30	1	246	5.82	1	2	0
104			3	8	35	9.7	3		YX7	Site1	8	1841	5.31	5	5	0
105			3	8	35	9.7	3		FD6	Site2	8	351	3.95	5	5	0
106			3	8	35	9.7	3		LY2	Site3	4	888	5.64	5	1	0
107			3	8	35	9.7	3		NE5	Site4	5	504	5.96	5	2	0
108			3	8	35	9.7	3		QZ8	Site5	6	1339	6.16	5	3	0
109			3	8	35	9.7	3		QQ2	Site6	5	746	5.43	4	2	0
110			3	8	35	9.7	3		ZW6	Site7	8	1514	2.84	4	5	1
111			3	8	35	9.7	3		BY9	Site8	7	1081	4.21	4	4	0
112			3	8	35	9.7	3		MM7	Site9	3	77	4.19	4	0	0
113			3	8	35	9.7	3		NB6	Site10	7	91	6.14	4	4	0
114			3	8	35	9.7	3		RQ5	Site11	6	155	5.06	3	3	0
115			3	8	35	9.7	3		QF9	Site12	5	55	6.16	3	2	0
116	5	Canada	3	8	35	9.7	3	26	ZW1	Site13	8	121	1.42	3	5	1
117			3	8	35	9.7	3		TE4	Site14	8	93	3.65	3	5	0
118			3	8	35	9.7	3		KE2	Site15	6	53	4.37	3	3	0
119			3	8	35	9.7	3		UN8	Site16	4	226	4.43	3	1	0
120			3	8	35	9.7	3		UV3	Site17	5	58	5.31	3	2	0
121			3	8	35	9.7	3		EU5	Site18	5	75	5.57	3	2	0
122			3	8	35	9.7	3		GG6	Site19	8	148	4.73	3	5	0
123			3	8	35	9.7	3		KS6	Site20	5	170	3.72	3	2	0
124			3	8	35	9.7	3		ZU4	Site21	6	179	4.79	2	3	0
125			3	8	35	9.7	3		VK5	Site22	3	208	5.93	1	0	0
126			3	8	35	9.7	3		OK2	Site23	5	61	3.90	2	2	0
127			3	8	35	9.7	3		IJ2	Site24	5	197	5.21	2	2	0
128			3	8	35	9.7	3		WV8	Site25	3	193	6.12	2	0	0
129			3	8	35	9.7	3		TD2	Site26	5	1500	4.91	1	2	0

130	6	Singapore	8	8	5	14	3	2.1	FO1	Site1	8	143	8.89	4	5	0
131			8	8	5	14	3		IA6	Site2	8	133	3.84	3	5	0
132	7	United Kingdom	0	0	64	10	3	28	XC2	Site1	0	1339	1.96	5	3	1
133			0	0	64	10	3		AJ8	Site2	0	1010	5.44	5	3	0
134			0	0	64	10	3		KY4	Site3	0	1551	5.27	5	3	0
135			0	0	64	10	3		LQ2	Site4	0	513	3.45	5	3	0
136			0	0	64	10	3		VQ6	Site5	0	927	3.17	5	3	0
137			0	0	64	10	3		ZM8	Site6	0	325	5.17	4	3	0
138			0	0	64	10	3		HS5	Site7	0	911	5.17	4	3	0
139			0	0	64	10	3		AA5	Site8	0	1330	3.28	4	3	0
140			0	0	64	10	3		GQ6	Site9	0	1060	5.24	4	3	0
141			0	0	64	10	3		BX3	Site10	0	225	6.29	4	3	0
142			0	0	64	10	3		WD2	Site11	0	98	5.38	4	3	0
143			0	0	64	10	3		QM9	Site12	0	126	2.97	3	3	1
144			0	0	64	10	3		WI2	Site13	0	157	4.76	3	3	0
145			0	0	64	10	3		PB8	Site14	0	114	5.92	3	3	0
146			0	0	64	10	3		UW3	Site15	0	155	6.06	3	3	0
147			0	0	64	10	3		DG7	Site16	0	92	5.25	3	3	0
148			0	0	64	10	3		OL4	Site17	0	77	3.07	3	3	0
149			0	0	64	10	3		XB9	Site18	0	127	1.41	3	3	1
150			0	0	64	10	3		FZ6	Site19	0	186	5.65	3	3	0
151			0	0	64	10	3		LG2	Site20	0	143	5.48	3	3	0
152			0	0	64	10	3		WU8	Site21	0	113	5.57	3	3	0
153			0	0	64	10	3		DQ6	Site22	0	128	4.28	3	3	0
154			0	0	64	10	3		VF9	Site23	0	237	4.13	3	3	0
155			0	0	64	10	3		AZ3	Site24	0	120	5.72	3	3	0
156			0	0	64	10	3		QZ9	Site25	0	132	6.06	2	3	0
157			0	0	64	10	3		VQ7	Site26	0	107	6.21	2	3	0
158			0	0	64	10	3		YD3	Site27	0	223	6.28	1	3	0
159			0	0	64	10	3		IX3	Site28	0	144	2.33	1	3	0
160			0	0	64	10	3		WO6	Site29	0	194	5.78	1	3	0
161			9	9	50	5	3		IK5	Site1	9	1283	0.91	5	6	1
162			9	9	50	5	3		CQ1	Site2	9	822	1.31	5	6	1
163			9	9	50	5	3		HU3	Site3	9	1144	2.56	5	6	1
164			9	9	50	5	3		BK1	Site4	9	422	3.11	5	6	0

165			9	9	50	5	3		RE3	Site5	9	909	2.86	4	6	1
166			9	9	50	5	3		GY3	Site6	9	287	2.48	4	6	1
167			9	9	50	5	3		UA4	Site7	9	209	2.78	4	6	1
168			9	9	50	5	3		ZN8	Site8	9	78	3.01	4	6	0
169	8	South Korea	9	9	50	5	3	18	XJ4	Site9	9	51	3.12	3	6	0
170			9	9	50	5	3		AK6	Site10	9	136	2.46	3	6	1
171			9	9	50	5	3		ZB7	Site11	9	152	2.67	3	6	1
172			9	9	50	5	3		LG7	Site12	9	84	2.61	3	6	1
173			9	9	50	5	3		LR3	Site13	9	82	2.79	3	6	1
174			9	9	50	5	3		BA5	Site14	9	119	1.11	3	6	1
175			9	9	50	5	3		HF7	Site15	9	172	1.99	3	6	1
176			9	9	50	5	3		LV3	Site16	9	148	2.71	3	6	1
177			9	9	50	5	3		QZ9	Site17	9	235	1.97	2	6	1
178			9	9	50	5	3		OY1	Site18	9	161	1.85	1	6	0
179			9	9	127	6.9	3		HD7	Site1	9	323	3.23	5	6	0
180			9	9	127	6.9	3		YV8	Site2	9	256	3.89	5	6	0
181			9	9	127	6.9	3		RK9	Site3	9	1061	2.80	5	6	1
182			9	9	127	6.9	3		JI1	Site4	9	1709	3.59	5	6	0
183			9	9	127	6.9	3		BE9	Site5	9	743	3.52	5	6	0
184			9	9	127	6.9	3		LS3	Site6	9	1702	3.77	5	6	0
185			9	9	127	6.9	3		JX2	Site7	9	675	3.18	4	6	0
186			9	9	127	6.9	3		XV3	Site8	9	1849	3.10	4	6	0
187			9	9	127	6.9	3		PQ1	Site9	9	100	4.21	4	6	0
188			9	9	127	6.9	3		GT5	Site10	9	95	3.44	4	6	0
189			9	9	127	6.9	3		PJ4	Site11	9	202	2.89	4	6	1
190			9	9	127	6.9	3		BS6	Site12	9	161	3.98	4	6	0
191			9	9	127	6.9	3		DR9	Site13	9	61	3.63	3	6	0
192			9	9	127	6.9	3		SO1	Site14	9	206	4.10	3	6	0
193	9	Japan	9	9	127	6.9	3	30.5	QS8	Site15	9	181	3.30	3	6	0
194			9	9	127	6.9	3		BG2	Site16	9	195	4.28	3	6	0
195			9	9	127	6.9	3		FF7	Site17	9	148	4.25	3	6	0
196			9	9	127	6.9	3		ON3	Site18	9	66	3.91	3	6	0
197			9	9	127	6.9	3		MZ4	Site19	9	211	3.97	3	6	0
198			9	9	127	6.9	3		DA9	Site20	9	74	4.19	3	6	0
199			9	9	127	6.9	3		ZT5	Site21	9	81	3.88	3	6	0

200			9	9	127	6.9	3		PH2	Site22	9	77	4.22	3	6	0
201			9	9	127	6.9	3		HR8	Site23	9	158	3.93	3	6	0
202			9	9	127	6.9	3		KT5	Site24	9	53	1.75	3	6	1
203			9	9	127	6.9	3		UY8	Site25	9	113	4.32	3	6	0
204			9	9	127	6.9	3		KX3	Site26	9	122	2.78	3	6	1
205			9	9	127	6.9	3		CP8	Site27	9	182	3.33	3	6	0
206			9	9	127	6.9	3		TH6	Site28	9	165	4.38	2	6	0
207			9	9	127	6.9	3		BM6	Site29	9	66	3.86	1	6	0
208			9	9	127	6.9	3		UW2	Site30	9	225	3.73	1	6	0
209	10	Finland	2	2	5	13	3	2.7	ML9	Site1	2	138	5.09	4	1	0
210			2	2	5	13	3		PJ7	Site2	2	248	3.89	3	1	0
211			8	8	200	3	2		QK3	Site1	8	944	1.38	5	5	1
212			8	8	200	3	2		KQ2	Site2	8	1808	1.86	5	5	1
213			8	8	200	3	2		WC5	Site3	8	1000	1.02	5	5	1
214			8	8	200	3	2		RB3	Site4	8	1601	1.38	5	5	1
215			8	8	200	3	2		TS3	Site5	8	987	0.72	4	5	1
216			8	8	200	3	2		XP4	Site6	8	646	0.74	4	5	1
217			8	8	200	3	2		HN5	Site7	8	1238	1.89	4	5	1
218			8	8	200	3	2		CK2	Site8	8	1710	1.88	4	5	1
219			8	8	200	3	2		PZ4	Site9	8	1822	0.85	4	5	1
220			8	8	200	3	2		GL4	Site10	8	260	1.38	4	5	1
221			8	8	200	3	2		OP2	Site11	8	1331	0.65	4	5	1
222			8	8	200	3	2		VY5	Site12	8	792	1.78	4	5	1
223			8	8	200	3	2		MU2	Site13	8	1259	1.01	4	5	1
224			8	8	200	3	2		AB8	Site14	8	1933	0.56	4	5	1
225			8	8	200	3	2		DL9	Site15	8	405	1.80	4	5	1
226			8	8	200	3	2		HE2	Site16	8	489	1.41	4	5	1
227			8	8	200	3	2		QZ9	Site17	8	80	1.53	4	5	1
228			8	8	200	3	2		JT2	Site18	8	219	1.75	4	5	1
229			8	8	200	3	2		XN3	Site19	8	188	0.71	4	5	1
230			8	8	200	3	2		EZ9	Site20	8	75	1.65	4	5	1
231			8	8	200	3	2		PQ5	Site21	8	197	0.79	4	5	1
232			8	8	200	3	2		KQ3	Site22	8	101	0.47	4	5	1
233			8	8	200	3	2		UQ8	Site23	8	172	1.74	4	5	1
234			8	8	200	3	2		LY3	Site24	8	54	1.87	4	5	1

235			8	8	200	3	2		VD8	Site25	8	135	0.48	4	5	1
236			8	8	200	3	2		PB8	Site26	8	87	1.13	4	5	1
237			8	8	200	3	2		SJ3	Site27	8	184	1.25	4	5	1
238			8	8	200	3	2		JN3	Site28	8	212	0.74	4	5	1
239			8	8	200	3	2		IN6	Site29	8	216	1.19	3	5	1
240			8	8	200	3	2		EK9	Site30	8	154	1.54	3	5	1
241			8	8	200	3	2		TX1	Site31	8	249	1.65	3	5	1
242			8	8	200	3	2		GT7	Site32	8	102	1.58	3	5	1
243			8	8	200	3	2		CO9	Site33	8	180	1.60	3	5	1
244			8	8	200	3	2		TX7	Site34	8	50	1.42	3	5	1
245			8	8	200	3	2		QO9	Site35	8	65	1.89	3	5	1
246			8	8	200	3	2		MH3	Site36	8	160	1.49	3	5	1
247			8	8	200	3	2		DX9	Site37	8	202	0.69	3	5	1
248			8	8	200	3	2		WY8	Site38	8	104	1.50	3	5	1
249			8	8	200	3	2		RU9	Site39	8	81	0.30	3	5	1
250			8	8	200	3	2		MA8	Site40	8	248	1.17	3	5	1
251	11	China	8	8	200	3	2	80	FB1	Site41	8	244	1.53	3	5	1
252			8	8	200	3	2		KM9	Site42	8	138	1.29	3	5	1
253			8	8	200	3	2		TT1	Site43	8	241	1.65	3	5	1
254			8	8	200	3	2		MI3	Site44	8	54	1.73	3	5	1
255			8	8	200	3	2		IU5	Site45	8	54	0.40	3	5	1
256			8	8	200	3	2		QY8	Site46	8	80	1.25	3	5	1
257			8	8	200	3	2		HU5	Site47	8	106	0.88	3	5	1
258			8	8	200	3	2		GT6	Site48	8	25	1.58	3	5	1
259			8	8	200	3	2		SB9	Site49	8	48	1.67	3	5	1
260			8	8	200	3	2		XN5	Site50	8	44	1.74	3	5	1
261			8	8	200	3	2		DA2	Site51	8	34	0.72	3	5	1
262			8	8	200	3	2		OU8	Site52	8	43	0.76	3	5	1
263			8	8	200	3	2		BM7	Site53	8	49	1.24	2	5	1
264			8	8	200	3	2		WZ3	Site54	8	30	1.82	2	5	1
265			8	8	200	3	2		HB6	Site55	8	30	0.81	2	5	1
266			8	8	200	3	2		AP9	Site56	8	28	0.82	2	5	1
267			8	8	200	3	2		LO6	Site57	8	25	1.79	2	5	1
268			8	8	200	3	2		KG2	Site58	8	26	1.77	2	5	1
269			8	8	200	3	2		RQ7	Site59	8	26	1.70	2	5	1

270			8	8	200	3	2		QC1	Site60	8	46	1.35	2	5	1
271			8	8	200	3	2		SG2	Site61	8	21	1.43	2	5	1
272			8	8	200	3	2		OM6	Site62	8	37	1.64	2	5	1
273			8	8	200	3	2		SQ8	Site63	8	23	1.85	2	5	1
274			8	8	200	3	2		SD6	Site64	8	23	1.78	2	5	1
275			8	8	200	3	2		VU3	Site65	8	37	1.99	2	5	1
276			8	8	200	3	2		CW6	Site66	8	38	1.81	2	5	1
277			8	8	200	3	2		EU3	Site67	8	35	1.37	2	5	1
278			8	8	200	3	2		WF3	Site68	8	21	1.95	2	5	1
279			8	8	200	3	2		VF4	Site69	8	37	1.40	2	5	1
280			8	8	200	3	2		JG7	Site70	8	43	0.12	2	5	1
281			8	8	200	3	2		BJ3	Site71	8	31	1.53	2	5	1
282			8	8	200	3	2		KO1	Site72	8	37	1.93	2	5	1
283			8	8	200	3	2		FM2	Site73	8	49	1.43	2	5	1
284			8	8	200	3	2		DV5	Site74	8	39	1.78	2	5	1
285			8	8	200	3	2		WF6	Site75	8	21	1.18	2	5	1
286			8	8	200	3	2		IC7	Site76	8	32	1.67	2	5	1
287			8	8	200	3	2		KR1	Site77	8	46	0.70	2	5	1
288			8	8	200	3	2		RT7	Site78	8	46	1.73	1	5	0
289			8	8	200	3	2		CR9	Site79	8	22	1.81	1	5	0
290			8	8	200	3	2		LS4	Site80	8	24	2.08	1	5	0
291			8	8	200	3	2		ET2	Site81	8	46	1.61	1	5	0
292			2	4	200	3	2		PH3	Site1	3	593	1.87	5	0	0
293			2	4	200	3	2		QZ8	Site2	4	1286	1.79	5	1	1
294			2	4	200	3	2		NH7	Site3	3	785	1.76	4	0	0
295			2	4	200	3	2		NN1	Site4	4	858	1.75	4	1	1
296			2	4	200	3	2		VE5	Site5	3	845	1.64	4	0	0
297			2	4	200	3	2		LU5	Site6	2	333	1.78	4	1	1
298			2	4	200	3	2		XK4	Site7	2	1179	1.44	4	1	1
299			2	4	200	3	2		SE3	Site8	4	73	0.16	4	1	1
300			2	4	200	3	2		LL1	Site9	4	57	0.65	4	1	1
301			2	4	200	3	2		BK3	Site10	2	59	1.45	4	1	1
302			2	4	200	2	2		VA9	Site11	4	136	0.70	4	1	1
303			2	4	200	2	2		PW4	Site12	2	57	0.60	4	1	1
304			2	4	200	2	2		UZ7	Site13	2	185	0.99	4	1	1

305			2	4	200	2	2		AZ7	Site14	4	84	1.14	3	1	1
306			2	4	200	2	2		MI7	Site15	2	189	1.32	3	1	1
307	12	Brazil	2	4	200	2	2	32	UO6	Site16	3	158	0.73	3	0	0
308			2	4	200	2	2		DT4	Site17	4	221	0.41	3	1	1
309			2	4	200	2	2		NH7	Site18	2	180	1.38	3	1	1
310			2	4	200	2	2		ZY7	Site19	3	198	0.56	3	0	0
311			2	4	200	2	2		KS5	Site20	2	225	1.28	3	1	1
312			2	4	200	2	2		JC5	Site21	3	155	1.17	3	0	0
313			2	4	200	2	2		PC3	Site22	3	54	1.00	3	0	0
314			2	4	200	2	2		DI1	Site23	3	23	1.19	3	0	0
315			2	4	200	2	2		TZ9	Site24	3	49	1.39	2	0	0
316			2	4	200	2	2		GK4	Site25	4	27	0.73	2	1	1
317			2	4	200	2	2		IH7	Site26	4	46	1.23	2	1	1
318			2	4	200	2	2		WF1	Site27	3	31	0.61	2	0	0
319			2	4	200	2	2		WT4	Site28	3	35	0.89	2	0	0
320			2	4	200	2	2		RU7	Site29	3	24	0.76	2	0	0
321			2	4	200	2	2		LF3	Site30	2	31	0.89	2	1	1
322			2	4	200	2	2		FO7	Site31	4	41	0.79	2	1	1
323			2	4	200	2	2		HH4	Site32	4	34	1.28	1	1	0
324			5.5	5.5	200	0.5	2		QC7	Site1	6	1871	0.11	5	3	1
325			5.5	5.5	200	0.5	2		EI4	Site2	6	549	0.21	5	3	1
326			5.5	5.5	200	0.5	2		KT1	Site3	6	1071	0.02	5	3	1
327			5.5	5.5	200	0.5	2		PQ3	Site4	6	1166	0.07	5	3	1
328			5.5	5.5	200	0.5	2		RW3	Site5	6	1751	0.33	5	3	1
329			5.5	5.5	200	0.5	2		YD4	Site6	6	319	0.31	4	3	1
330			5.5	5.5	200	0.5	2		DX4	Site7	6	1546	0.05	4	3	1
331			5.5	5.5	200	0.5	2		MV7	Site8	6	1388	0.28	4	3	1
332			5.5	5.5	200	0.5	2		TD7	Site9	6	652	0.29	4	3	1
333			5.5	5.5	200	0.5	2		CL6	Site10	6	887	0.33	4	3	1
334			5.5	5.5	200	0.5	2		SH9	Site11	6	1169	0.24	4	3	1
335			5.5	5.5	200	0.5	2		QI5	Site12	6	1783	0.20	4	3	1
336			5.5	5.5	200	0.5	2		JL4	Site13	6	253	0.30	4	3	1
337			5.5	5.5	200	0.5	2		VD5	Site14	6	1567	0.06	4	3	1
338			5.5	5.5	200	0.5	2		VC9	Site15	6	1543	0.30	4	3	1
339			5.5	5.5	200	0.5	2		LF6	Site16	6	1469	0.33	4	3	1

340			5.5	5.5	200	0.5	2		JL3	Site17	6	1560	0.16	4	3	1
341			5.5	5.5	200	0.5	2		LA6	Site18	6	234	0.15	4	3	1
342			5.5	5.5	200	0.5	2		DP1	Site19	6	86	0.15	4	3	1
343			5.5	5.5	200	0.5	2		SH1	Site20	6	67	0.33	4	3	1
344			5.5	5.5	200	0.5	2		KA9	Site21	6	87	0.25	4	3	1
345			5.5	5.5	200	0.5	2		DT7	Site22	6	99	0.33	4	3	1
346			5.5	5.5	200	0.5	2		JR1	Site23	6	134	0.31	4	3	1
347			5.5	5.5	200	0.5	2		NZ1	Site24	6	219	0.16	4	3	1
348			5.5	5.5	200	0.5	2		UG1	Site25	6	186	0.32	4	3	1
349			5.5	5.5	200	0.5	2		RF4	Site26	6	182	0.29	4	3	1
350			5.5	5.5	200	0.5	2		VP5	Site27	6	143	0.34	4	3	1
351			5.5	5.5	200	0.5	2		BP2	Site28	6	226	0.32	4	3	1
352			5.5	5.5	200	0.5	2		MW1	Site29	6	85	0.31	4	3	1
353			5.5	5.5	200	0.5	2		JH4	Site30	6	128	0.10	4	3	1
354			5.5	5.5	200	0.5	2		WP4	Site31	6	212	0.15	4	3	1
355			5.5	5.5	200	0.5	2		YO1	Site32	6	238	0.31	4	3	1
356			5.5	5.5	200	0.5	2		MK1	Site33	6	53	0.34	4	3	1
357			5.5	5.5	200	0.5	2		OS1	Site34	6	99	0.20	4	3	1
358			5.5	5.5	200	0.5	2		QV1	Site35	6	59	0.15	4	3	1
359			5.5	5.5	200	0.5	2		KZ3	Site36	6	198	0.35	3	3	1
360			5.5	5.5	200	0.5	2		NY5	Site37	6	133	0.18	3	3	1
361			5.5	5.5	200	0.5	2		EB5	Site38	6	110	0.11	3	3	1
362			5.5	5.5	200	0.5	2		YR7	Site39	6	194	0.07	3	3	1
363			5.5	5.5	200	0.5	2		ON8	Site40	6	120	0.28	3	3	1
364			5.5	5.5	200	0.5	2		LH4	Site41	6	139	0.31	3	3	1
365			5.5	5.5	200	0.5	2		KD8	Site42	6	68	0.06	3	3	1
366			5.5	5.5	200	0.5	2		BT3	Site43	6	125	0.08	3	3	1
367			5.5	5.5	200	0.5	2		QR7	Site44	6	232	0.05	3	3	1
368			5.5	5.5	200	0.5	2		GJ6	Site45	6	53	0.10	3	3	1
369			5.5	5.5	200	0.5	2		ZV3	Site46	6	64	0.29	3	3	1
370			5.5	5.5	200	0.5	2		TF2	Site47	6	164	0.14	3	3	1
371			5.5	5.5	200	0.5	2		BX4	Site48	6	102	0.15	3	3	1
372			5.5	5.5	200	0.5	2		FS1	Site49	6	249	0.28	3	3	1
373	13	India	5.5	5.5	200	0.5	2	100	OW9	Site50	6	143	0.29	3	3	1
374			5.5	5.5	200	0.5	2		FE2	Site51	6	79	0.15	3	3	1

375			5.5	5.5	200	0.5	2		VU2	Site52	6	58	0.22	3	3	1
376			5.5	5.5	200	0.5	2		KL7	Site53	6	100	0.15	3	3	1
377			5.5	5.5	200	0.5	2		YL2	Site54	6	230	0.32	3	3	1
378			5.5	5.5	200	0.5	2		BK4	Site55	6	165	0.26	3	3	1
379			5.5	5.5	200	0.5	2		VG4	Site56	6	179	0.31	3	3	1
380			5.5	5.5	200	0.5	2		HD5	Site57	6	133	0.22	3	3	1
381			5.5	5.5	200	0.5	2		PE2	Site58	6	190	0.31	3	3	1
382			5.5	5.5	200	0.5	2		NA1	Site59	6	129	0.05	3	3	1
383			5.5	5.5	200	0.5	2		RC6	Site60	6	58	0.32	3	3	1
384			5.5	5.5	200	0.5	2		YL7	Site61	6	198	0.22	3	3	1
385			5.5	5.5	200	0.5	2		OK1	Site62	6	161	0.18	3	3	1
386			5.5	5.5	200	0.5	2		DW5	Site63	6	186	0.31	3	3	1
387			5.5	5.5	200	0.5	2		OG7	Site64	6	148	0.22	3	3	1
388			5.5	5.5	200	0.5	2		XH7	Site65	6	178	0.34	2	3	1
389			5.5	5.5	200	0.5	2		IO6	Site66	6	70	0.24	2	3	1
390			5.5	5.5	200	0.5	2		UQ1	Site67	6	197	0.04	2	3	1
391			5.5	5.5	200	0.5	2		BH9	Site68	6	190	0.09	2	3	1
392			5.5	5.5	200	0.5	2		OP9	Site69	6	70	0.29	2	3	1
393			5.5	5.5	200	0.5	2		OV6	Site70	6	158	0.22	2	3	1
394			5.5	5.5	200	0.5	2		OM7	Site71	6	94	0.19	2	3	1
395			5.5	5.5	200	0.5	2		TQ8	Site72	6	165	0.27	2	3	1
396			5.5	5.5	200	0.5	2		FZ6	Site73	6	61	0.13	2	3	1
397			5.5	5.5	200	0.5	2		WP1	Site74	6	173	0.27	2	3	1
398			5.5	5.5	200	0.5	2		AO8	Site75	6	102	0.19	2	3	1
399			5.5	5.5	200	0.5	2		JT2	Site76	6	174	0.34	2	3	1
400			5.5	5.5	200	0.5	2		NP7	Site77	6	126	0.15	2	3	1
401			5.5	5.5	200	0.5	2		OV1	Site78	6	83	0.17	2	3	1
402			5.5	5.5	200	0.5	2		MJ4	Site79	6	139	0.29	2	3	1
403			5.5	5.5	200	0.5	2		DF5	Site80	6	71	0.15	2	3	1
404			5.5	5.5	200	0.5	2		FV1	Site81	6	66	0.31	2	3	1
405			5.5	5.5	200	0.5	2		RK2	Site82	6	242	0.13	2	3	1
406			5.5	5.5	200	0.5	2		JY5	Site83	6	31	0.25	2	3	1
407			5.5	5.5	200	0.5	2		NS4	Site84	6	35	0.34	2	3	1
408			5.5	5.5	200	0.5	2		WM1	Site85	6	41	0.19	2	3	1
409			5.5	5.5	200	0.5	2		JK5	Site86	6	22	0.06	2	3	1

410			5.5	5.5	200	0.5	2		YC2	Site87	6	33	0.01	2	3	1
411			5.5	5.5	200	0.5	2		FS7	Site88	6	42	0.28	2	3	1
412			5.5	5.5	200	0.5	2		HC2	Site89	6	24	0.15	2	3	1
413			5.5	5.5	200	0.5	2		PA8	Site90	6	28	0.24	2	3	1
414			5.5	5.5	200	0.5	2		HP7	Site91	6	31	0.26	2	3	1
415			5.5	5.5	200	0.5	2		EH9	Site92	6	46	0.17	2	3	1
416			5.5	5.5	200	0.5	2		XY4	Site93	6	49	0.24	2	3	1
417			5.5	5.5	200	0.5	2		AV2	Site94	6	35	0.07	1	3	0
418			5.5	5.5	200	0.5	2		AG9	Site95	6	48	0.11	1	3	0
419			5.5	5.5	200	0.5	2		NM8	Site96	6	49	0.24	1	3	0
420			5.5	5.5	200	0.5	2		SM6	Site97	6	40	0.15	1	3	0
421			5.5	5.5	200	0.5	2		EL6	Site98	6	45	0.29	1	3	0
422			5.5	5.5	200	0.5	2		YP2	Site99	6	32	0.16	1	3	0
423			5.5	5.5	200	0.5	2		FU4	Site100	6	23	0.19	1	3	0
424	14	Jordan	2	2	6.4	1.29	2	2.3	CW5	Site1	2	198	0.80	3	1	1
425			2	2	6.4	1.29	2		KA6	Site2	2	100	0.19	2	1	1
426	15	Egypt	2	2	82	1	2	29.5	HU8	Site1	2	1385	0.42	5	1	1
427			2	2	82	1	2		EA4	Site2	2	481	0.59	5	1	1
428			2	2	82	1	2		DW7	Site3	2	429	0.62	4	1	1
429			2	2	82	1	2		GM6	Site4	2	1861	0.47	4	1	1
430			2	2	82	1	2		KB8	Site5	2	796	0.53	4	1	1
431			2	2	82	1	2		PU1	Site6	2	991	0.36	4	1	1
432			2	2	82	1	2		OB8	Site7	2	638	0.48	4	1	1
433			2	2	82	1	2		MT9	Site8	2	235	0.16	4	1	1
434			2	2	82	1	2		YI6	Site9	2	53	0.17	4	1	1
435			2	2	82	1	2		GX8	Site10	2	162	0.57	3	1	1
436			2	2	82	1	2		AA1	Site11	2	182	0.45	3	1	1
437			2	2	82	1	2		AL5	Site12	2	186	0.45	3	1	1
438			2	2	82	1	2		HU3	Site13	2	238	0.33	3	1	1
439			2	2	82	1	2		DU7	Site14	2	67	0.17	3	1	1
440			2	2	82	1	2		MH8	Site15	2	159	0.47	3	1	1
441			2	2	82	1	2		IT8	Site16	2	250	0.19	3	1	1
442			2	2	82	1	2		DM1	Site17	2	228	0.51	2	1	1
443			2	2	82	1	2		OF8	Site18	2	53	0.07	2	1	1
444			2	2	82	1	2		MM3	Site19	2	92	0.33	2	1	1

445			2	2	82	1	2		UY8	Site20	2	206	0.43	2	1	1
446			2	2	82	1	2		HR6	Site21	2	42	0.30	2	1	1
447			2	2	82	1	2		MF3	Site22	2	44	0.45	2	1	1
448			2	2	82	1	2		SI6	Site23	2	39	0.30	2	1	1
449			2	2	82	1	2		UR9	Site24	2	42	0.12	2	1	1
450			2	2	82	1	2		LG1	Site25	2	34	0.33	1	1	0
451			2	2	82	1	2		TS4	Site26	2	23	0.48	1	1	0
452		Turkey	2	2	8	3.2	2	2	CD3	Site1	2	140	1.99	3	1	1
453			2	2	74	3.2	2		KI7	Site2	2	231	1.43	4	1	1
454	17	Ukraine	2	2	74	1	2	8	JP2	Site1	2	1139	0.42	5	1	1
455			2	2	74	1	2		VE1	Site2	2	1051	0.58	4	1	1
456			2	2	74	1	2		PI7	Site3	2	1231	0.34	4	1	1
457			2	2	74	1	2		FL8	Site4	2	246	0.54	3	1	1
458			2	2	74	1	2		CV4	Site5	2	240	0.47	3	1	1
459			2	2	74	1	2		NK7	Site6	2	49	0.62	3	1	1
460			2	2	74	1	2		SV8	Site7	2	50	0.21	2	1	1
461			2	2	45	1	2		FF4	Site8	2	25	0.57	1	1	0
462	19	South Africa	2	2	52	0.7	2	6	IG2	Site1	2	90	0.02	4	1	1
463			2	2	52	0.7	2		KO2	Site2	2	82	0.35	4	1	1
464			2	2	52	0.7	2		VA5	Site3	2	81	0.40	3	1	1
465			2	2	52	0.7	2		WK6	Site4	2	28	0.22	3	1	1
466			2	2	52	0.7	2		MS4	Site5	2	46	0.41	2	1	1
467			2	2	52	0.7	2		IG7	Site6	2	42	0.37	2	1	1
468	20	Thailand	7	7	67	1.2	2	21.4	JQ2	Site1	7	208	0.67	5	4	1
469			7	7	67	1.2	2		CP3	Site2	7	233	0.71	4	4	1
470			7	7	67	1.2	2		WQ7	Site3	7	173	0.70	4	4	1
471			7	7	67	1.2	2		QI3	Site4	7	223	0.05	4	4	1
472			7	7	67	1.2	2		ZT7	Site5	7	104	0.41	4	4	1
473			7	7	67	1.2	2		QX3	Site6	7	140	0.20	4	4	1
474			7	7	67	1.2	2		EV3	Site7	7	53	0.51	4	4	1
475			7	7	67	1.2	2		MQ2	Site8	7	101	0.55	3	4	1
476			7	7	67	1.2	2		SO4	Site9	7	234	0.24	3	4	1
477			7	7	67	1.2	2		XU3	Site10	7	161	0.75	3	4	1
478			7	7	67	1.2	2		UZ4	Site11	7	136	0.36	3	4	1
479			7	7	67	1.2	2		TT2	Site12	7	122	0.40	3	4	1

480			7	7	67	1.2	2		RT9	Site13	7	56	0.75	3	4	1
481			7	7	67	1.2	2		LV4	Site14	7	168	0.32	2	4	1
482			7	7	67	1.2	2		XM6	Site15	7	194	0.74	2	4	1
483			7	7	67	1.2	2		PQ9	Site16	7	94	0.05	2	4	1
484			7	7	67	1.2	2		PZ5	Site17	7	26	0.45	2	4	1
485			7	7	67	1.2	2		DW5	Site18	7	43	0.69	2	4	1
486			7	7	67	1.2	2		GP5	Site19	7	46	0.57	2	4	1
487			7	7	67	1.2	2		HD1	Site20	7	38	0.70	1	4	0
488			7	7	67	1.2	2		UE6	Site21	7	40	0.62	1	4	0
489	21	Zambia	2	2	14	0.47	1	1.4	TD1	Site1	2	35	0.05	3	1	1
490	22	Chad	1	1	12	0.7	1	1.68	MS2	Site1	1	25	0.33	3	2	1
491	23	Ethiopia	3	3	94	0.4	1	2	SQ3	Site1	3	36	0.14	3	0	0
492			3	3	94	0.4	1		AX8	Site2	3	47	0.09	2	0	0
493	24	Sudan	3	3	37	0.52	1	2.22	HB1	Site1	3	47	0.13	3	0	0
494			3	3	37	0.52	1		ZM2	Site2	3	29	0.11	2	0	0
495	25	Tanzania	3	3	49	0.13	1	2	EE5	Site1	3	40	0.02	3	0	0
496			3	3	49	0.13	1		XW7	Site2	3	48	0.05	2	0	0
497	26	Niger	1	1	17	0.35	1	2.38	NS8	Site1	1	42	0.03	3	2	1
498			1	1	17	0.35	1		NS3	Site2	1	49	0.17	2	2	1
499	27	Cambodia	7	7	15	0.3	1	2.1	AT2	Site1	7	46	0.18	2	4	1
500			7	7	15	0.3	1		YG7	Site2	7	41	0.11	3	4	1
501	28	Laos	7	7	6	0.3	1	0.6	PF6	Site1	7	27	0.18	3	4	1
502	29	Yemen	3	3	24	0.5	1	2	IG3	Site1	3	33	0.08	3	0	0
503			3	3	24	0.5	1		ZW5	Site2	3	49	0.12	2	0	0
504	30	Afghanistan	4.5	4.5	30	0.5	1	2	SJ1	Site1	5	32	0.29	3	2	1
505			4.5	4.5	30	0.5	1		WH3	Site2	5	40	0.13	2	2	1

6.5 VIATE

Vitae

- Name: Abdulrhman Ahmed Mohsen Alsri.
- Nationality: Yemeni.
- Date of Birth: 28/05/1986.
- Email: *man9500@gmail.com.*
- Permenant Address: Yemen, Hadhramout, Tareem.
- Publications: Abdulrhman Alsri, Sultan Almuhammadi and Sajjad Mahmood "A Model for Work Distribution in Global Software Development Based on Machine Learning Techniques", In Proceedings of the Science and Information Conference (SAI 2014), pp. 399 - 403, IEEE Computer Society 2014.